

Systematic analysis of heterochromatin modification readout

Dissertation

to acquire the doctoral degree in mathematics and natural science

'Doctor rerum naturalium'

at the Georg-August-Universität Göttingen

in the doctoral degree programme Biomolecules: Structure – Function - Dynamics

at the Georg-August University

School of Science (GAUSS)

Submitted by

Nadin Zimmermann

born in

Berlin, Germany

Göttingen, 2016

Thesis Committee

Dr. Wolfgang Fischle, Chromatin Biochemistry, Max Planck Institute for Biophysical Chemistry

Prof. Dr. Holger Stark, Dep. of Structural Dynamics, Max Planck Institute for Biophysical Chemistry

Prof. Dr. André Fischer, Dep. for Psychiatry and Psychotherapy, University Medical Center, German Center for Neurodegenerative Diseases

Members of the examination board:

Referee: Dr. Wolfgang Fischle, Chromatin Biochemistry, Max Planck Institute for Biophysical Chemistry

Co-referee: Prof. Dr. Holger Stark, Dep. of Structural Dynamics, Max Planck Institute for Biophysical Chemistry

Other members of the Examination Board:

Prof. Dr. André Fischer, Dep. for Psychiatry and Psychotherapy, University Medical Center, German Center for Neurodegenerative Diseases

Prof. Dr. Henning Urlaub, Research Group Mass Spectrometry, Max Planck Institute for Biophysical Chemistry

Prof. Dr. Detlef Doenecke, Dept. of Molecular Biology, University Medical Center Göttingen

Prof. Dr. Steven Johnsen, Clinic for General, Visceral and Pediatric Surgery, University Medical Center Göttingen

Date of the oral examination: 15th June 2016

Affidavit

I hereby declare that the presented thesis "Systematic analysis of heterochromatin modification readout" has been written independently and with no other sources and aids than quoted.

Göttingen, 10th May 2016

Nadin Zimmermann

Acknowledgements

First and foremost, I would like to express my gratitude to my mentors, Dr. Wolfgang Fischle and Prof. Dr. Henning Urlaub. Thank you for entrusting me this project, for your continuous support, guidance and your enthusiasm.

I am grateful to the members of my thesis committee – Prof. Dr. Holger Stark and Prof. Dr. André Fischer for their interest in the project, the helpful discussions and suggestions. I also thank Prof. Dr. Detlef Doenecke and Prof. Dr. Steven Johnsen for their time as part of my thesis examination committee.

I want to thank the GGNB for their generous financial support, informative lectures, highly supportive method courses as well as inspiring retreats.

My sincere thanks to all the present and past members of the Chromatin Biochemistry group for providing a great working atmosphere. I particularly thank my supervisor Miro sharing all his knowledge about the project, for all the difficult discussions, his support, his patience, and his trust. I am grateful to Alexandra for sharing her knowledge of protein biochemistry and for never running out of answers to all my questions. I thank Kyoko for all the fruitful discussions. I am especially grateful to Alexandra, Henriette und Maria for their friendship and the valuable hours spend outside of the lab.

I am very grateful to all the members of the Mass Spectrometry group for their support during the last four years. I want to thank Henning and Christof, for their support, the many fruitful cigarette-discussions, for Esprit, barbeques and many more. I also thank Uwe, Moni and Annika for their help and the introduction to the mass spectrometers.

I am thankful to Thomas Conrad for providing and culturing HeLa cells for me.

I thank the people from the climbing gym for giving me good reasons to leave the lab and the opportunity to clear my head. I also want to thank all my friends supporting me from far away over all the years.

I am very thankful to Cyrille for his support, the helpful discussions, his patience and all the chocolate he provided the last months.

I owe my gratitude to my family. In particular I want to thank my parents and my sister for constantly supporting and encouraging me to go all the way and for cheering me up during difficult times.

Table of contents

Table of contents.....	I
List of figures.....	VI
List of tables.....	VII
Abbreviations	VIII
Abstract.....	1
1 Introduction	2
1.1 Chromatin	2
1.1.1 Chromatin organization	2
1.1.2 Chromatin modifications.....	4
1.1.3 Readout of chemical chromatin modifications	8
1.1.4 Multivalent readout and crosstalk of chemical chromatin modifications	11
1.1.5 Chromatin arrays for systematic analysis of chemical modification readout.....	13
1.2 Mass spectrometry	14
1.2.1 Mass spectrometry based protein identification of complex samples.....	14
1.2.2 Relative quantification by Stable Isotope Labeling by Amino acids in Cell culture (SILAC)	16
1.2.3 Protein-protein cross-linking mass spectrometry	17
1.3 Hypothesis and objectives of the presented thesis	19
2 Materials and Methods.....	20
2.1 Material and reagents.....	20
2.1.1 Laboratory equipment	20
2.1.2 Consumables and plastic ware.....	21
2.1.3 Chemicals	21
2.1.4 Chromatographic and affinity material	23
2.1.5 Cell culture media and materials	23
2.1.6 Commercial kits.....	23
2.1.7 Commonly used buffers and solutions.....	24

2.1.8	Cell lines	27
2.1.9	Chemically competent <i>Escherichia coli</i> strains	28
2.1.10	Plasmids	28
2.1.11	Enzymes, Proteins and inhibitors	28
2.1.12	Antibodies	29
2.1.13	Antibiotics.....	29
2.1.14	Oligonucleotides.....	29
2.1.15	Peptides	30
2.1.16	Software.....	30
2.2	Molecular biology methods.....	31
2.2.1	Transformation of chemically competent bacteria.....	31
2.2.2	Analysis of nucleic acids	31
2.2.2.1	Concentration determination of nucleic acids	31
2.2.2.2	Nucleic acid gel electrophoresis	31
2.2.3	DNA precipitation	32
2.2.3.1	Ethanol.....	32
2.2.3.2	Isopropanol	32
2.2.3.3	PEG-6000	32
2.2.4	Preparation of DNA templates for chromatin array assembly	32
2.2.4.1	Preparation of biotinylated 12x200-601 DNA template	32
2.2.4.2	Methylation of biotinylated 12x200-601DNA template	33
2.2.4.3	Preparation of scavenger DNA.....	33
2.3	Protein biochemistry methods	34
2.3.1	Detection and analysis of proteins.....	34
2.3.1.1	Concentration determination of proteins	34
2.3.1.2	Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).....	35
2.3.1.3	Coomassie staining of protein SDS-PAGE gels.....	35
2.3.1.4	Silver staining of protein SDS-PAGE gels	35
2.3.1.5	Western blot.....	36
2.3.2	Recombinant proteins	36
2.3.2.1	Expression and purification of recombinant histone proteins	36
2.3.2.2	Introduction of posttranslational modifications by native chemical ligation	37

2.3.2.3	Introduction of methyl lysine analogs.....	38
2.3.3	Preparation of recombinant chromatin.....	38
2.3.3.1	Assembly of histone octamers.....	38
2.3.3.2	Chromatin reconstitution by salt dialysis.....	39
2.3.4	Molecular characterization of recombinant chromatin.....	39
2.3.4.1	Aval digest	39
2.3.4.2	Analytical ultracentrifugation.....	40
2.4	Cell culture and metabolic labeling	40
2.4.1	Stable isotope labeling by amino acids in cell culture (SILAC) of HeLa S3 cells ...	40
2.4.2	Preparation of nuclear extracts.....	40
2.5	Biochemical binding assays	41
2.5.1	Chromatin affinity purification	41
2.5.2	Protein-protein cross-linking on beads	42
2.6	Mass spectrometry methods	42
2.6.1	In-gel proteolysis of proteins and peptide extraction	42
2.6.2	In-solution proteolysis of proteins	43
2.6.3	LC-MS/MS analysis of peptides on LTQ-Orbitrap.....	43
2.6.4	LC-MS/MS analysis of peptides on QExactive.....	44
2.6.5	Molecular weight determination of histone proteins	45
2.7	Raw data processing and data analysis	45
2.7.1	Mass spectrometric raw data processing with MaxQuant	45
2.7.2	Mass spectrometric raw data processing with pLink	45
2.7.3	Data filtering and visualization of quantified MS data.....	46
2.7.4	Protein-protein interaction analysis.....	47
2.7.5	Annotation enrichment analysis.....	47
3	Results.....	48
3.1	12mer nucleosomal arrays are of reproducible quality independent of incorporated modifications.....	48
3.1.1	Native chemical ligation and alkylation of cysteine residues successfully introduce PTMs to recombinant <i>Xenopus laevis</i> histones	48

3.1.2	Octamer-DNA molar ratios from 1.1 to 1.3 assemble fully saturated, reproducible nucleosomal arrays	50
3.2	The statistical tool “significance A” determines significant protein fold enrichment cutoffs	50
3.3	ChAP coupled to quantitative MS enables the investigation of chromatin-protein interactomics and provides insights into chromatin modification crosstalk.....	55
3.3.1	Individual chromatin modification states recruit specific chromatin-binding interactomes.....	56
3.3.2	The methylation degrees of lysine residues affect protein binding to different extends	72
3.3.3	Two chromatin modifications <i>in trans</i> do not only constitute the sum of their single counterparts but demonstrate an independent impact on chromatin-binding interactomes .	74
3.3.3.1	Chromatin arrays carrying a combination of H3K9me3 and H4K20me3 result in a specific protein interactome.....	75
3.3.3.2	The double modification H3K9me3 H4K20me3 indicates positive and negative crosstalk	76
3.3.3.3	Communication of histone and DNA modifications is indicated by positive and negative crosstalk	79
3.4	Annotation analysis functionally correlates chromatin marks	84
3.5	Protein-protein cross-linking of the chromatin-binding interactome.....	88
3.5.1	Protein-protein cross-linking on assembled chromatin arrays.....	88
3.5.2	Protein-protein cross-links of complex protein samples are specifically detected by the search algorithm pLink in combination with a reference database	93
3.6	ChAP-MS coupled with XL provides information of the binding hierarchy beyond primary binding proteins recruited to chromatin.....	95
4	Discussion	99
4.1	Individual chromatin-binding interactomes are comparable based on significant enrichment cutoffs	99
4.2	Interactomics of chromatin-binding proteins provide new insights into the functional impact of individual chromatin modifications	100
4.3	Protein-binding interactomes of modified chromatin arrays display only moderate overlap compared to those revealed with different modified templates	105
4.4	Individual heterochromatic marks display positive as well as negative functional correlation.....	106

4.5	Combinatorial readout <i>in trans</i> provides new insights into crosstalk between chemical chromatin modifications	107
4.6	Protein-protein cross-linking coupled to MS maps the binding hierarchy of chromatin-bound proteins	109
5	Conclusions and future perspectives.....	111
6	References	112
7	Appendix.....	126
8	Curriculum Vitae	138

List of figures

Figure 1.1 Structure of the nucleosome core particle.....	3
Figure 1.2 Combinatorial readout of posttranslational histone modifications.....	12
Figure 1.3 Design of recombinant nucleosomal arrays.	14
Figure 1.4 Workflow for protein identification by MS.	15
Figure 3.1 Introduction of posttranslational histone modifications.	49
Figure 3.2 Quality controls of assembled chromatin arrays.	51
Figure 3.3 Workflow of SILAC-based chromatin affinity purification coupled with mass spectrometry.	52
Figure 3.4 Determination and visualization of significantly regulated proteins using the example of the H3K9me3 interactome.....	54
Figure 3.5 Protein-binding interactomes of H3K9me1 and H3K9me2 chromatin.....	57
Figure 3.6 Protein-binding interactome of H3K9me3-modified chromatin.	62
Figure 3.7 Protein-binding interactomes of H3K _C 27 mono- and di-methylated chromatin.	65
Figure 3.8 H3K _C 27me3 chromatin-binding interactome.	66
Figure 3.9 Chromatin-associated protein interactomes of histone H4K20 mono- and trimethylation.	68
Figure 3.10 Chromatin-binding proteins regulated by H4R3me2 and H3Δ1-20.....	70
Figure 3.11 Protein-binding interactome of chromatin containing methylated DNA.	71
Figure 3.12 Overlap of proteins regulated by the methylation degrees of H3K9.....	72
Figure 3.13 Overlap of significant proteins regulated by different methylation states.	74
Figure 3.14 Protein-binding interactome of H3K9me3 H4K20me3-modified chromatin.....	76
Figure 3.15 Comparison of significant protein binding to chromatin affected by H3K9me3, H4K20me3 and H3K9me3 H4K20me3.....	78
Figure 3.16 Crosstalk between H3K9me3 and H4K20me3.....	80
Figure 3.17 Chromatin-binding interactome of H3K9me3 meCpG in comparison to the individual modifications.....	81
Figure 3.18 Crosstalk of H3K9me3 and CpG-methylated DNA.....	83
Figure 3.19 Annotation enrichment analyses of biological pathways and GO terms of molecular functions.	86
Figure 3.20 General workflow of protein-protein cross-link coupled with ChAP-MS.	89
Figure 3.21 Quality control of reconstituted nucleosomal arrays.....	90
Figure 3.22 Enrichment for protein-protein cross-links by size exclusion chromatography (SEC).	91
Figure 3.23 Chemical cross-links of peptides using the cross-linker BS3.....	92
Figure 3.24 Protein-protein cross-links in the context of three-dimensional protein structures. ...	94

Figure 3.25 Protein-protein cross-link analysis of the core nucleosome.....	96
Figure 3.26 Interaction network of cross-linked factors associated with chromatin.	98

List of tables

Table 1.1 Histone modifications.....	7
Table 2.1 Plasmids used for chromatin reconstitution and protein expression.	28
Table 2.2 Antibodies and antisera used for western blot analysis.	29
Table 2.3 Peptides used for native chemical ligation.	30
Table 2.4 Values used for the determination of histone concentrations.	34
Table 3.1 Proteins recruited and excluded from binding to chemically modified chromatin.	59

Abbreviations

aa	amino acid(s)	min	minute(s)
ACN	acetonitrile	MLA	methyl lysine analog
APS	ammonium peroxodisulfate	MPAA	4-Mercaptophenylacetic acid
ATP	adenosine triphosphate	MS	mass spectrometry
bp	base pairs	MS/MS	tandem mass spectrometry
BS3	bis(sulfosuccinimidyl) suberate	Mw	molecular weight
BSA	bovine serum albumin	MWC	molecular weight cut-off
C	Celsius	NCL	native chemical ligation
ChAP	chromatin affinity purification	NE	nuclear extract
ChIP	chromatin immunoprecipitation	OD	optical density
Da	Dalton (g/mol)	PAGE	polyacrylamide gel electrophoresis
DDA	data-dependent acquisition	PBS	phosphate buffered saline
ddH ₂ O	double distilled water (sterilized)	PCI	phenol:chloroform:isoamyl alcohol
DMEM	Dulbecco's Modified Eagle's medium	PEG	polyethylene glycol
DNA	deoxyribonucleic acid	PMSF	phenylmethylsulfonyl fluoride
ds	double-stranded	ppm	parts per million
DTT	dithiothreitol	PTM	posttranslational modification
<i>E.coli</i>	<i>Escherichia coli</i>	RNA	ribonucleic acid
e.g.	for example, exempli gratia	RP	reversed phase
EDTA	ethylenediaminetetraacetic acid	rpm	revolutions per minute
ESI	electrospray ionization	Rt	room temperature
et al.	and others, et alii	s	second(s)
FA	formic acid	SDS	sodium dodecyl sulfate
FBS	fetal bovine serum	SEC	size exclusion chromatography
FW HM	full width at half maximum	SILAC	stable isotope labeling by amino acids in cell culture
g	gram or gravity force	TBE	tris/borate/EDTA
h	hour	TEMED	N,N,N',N'-tetramethylethylenediamine
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid	Tris	tris-(hydroxymethyl) aminomethane
HPLC	high pressure liquid chromatography	TSS	transcription start site
IAA	iodoacetamide	UV	ultraviolet
k	kilo	V	volt
KDM	lysine demethylase	v/v	volume per volume
KMT	lysine methyltransferase	w/v	weight per volume
l	liter	WT	wild type
LC	liquid chromatography	<i>X. laevis</i>	<i>Xenopus laevis</i>
LTQ	linear trap quadrupole	XL	cross-link
m	milli or meter	α	anti-/antibody
M	molar	Δ	delta (deletion)
<i>m/z</i>	mass-to-charge ratio	μ	micro

Abstract

Modifications of the building units of chromatin - the histone proteins and DNA - direct the functional readout of the genome and are part of the so called “epigenome” (Fischle et al., 2003b; Kouzarides, 2007). Several studies have shown that specialized proteins recognize these chemical modifications and thereby define the functional state of chromatin (Bartke et al., 2010; Nikolov et al., 2011; Vermeulen et al., 2010). Several individual chromatin-binding factors have been analyzed in detail. Also, proteomics approaches have taken inventory of factors interacting with single chromatin modifications. However, the full set of factors specific to particular modified chromatin domains, such as for example eu- and heterochromatin, is not known yet.

In this thesis, I report my findings from a systematic *in vitro* chromatin affinity purification approach in combination with quantitative mass spectrometry (Nikolov et al., 2011) that significantly expands the list of proteins regulated by heterochromatic modifications, both alone and in combination. The results provide a comprehensive catalogue of proteins specifically regulated in their chromatin binding by specific modification patterns. Moreover, the datasets highlight novel biological functions associated with individual modifications and reveal unpredicted functional relationships of the investigated chromatin modifications.

The analysis of the interactomes of combinatorially modified chromatin identifies for the first time proteins whose binding properties to chromatin depend on the cooperative action of two modifications. The results provide novel insights into the communication between two chromatin modifications, namely positive and negative crosstalk.

Furthermore, I introduce a new experimental workflow combining chromatin affinity purification and cross-linking mass spectrometry. This workflow maps physical protein-protein interactions sites of chromatin-bound proteins and provides first insights into the hierarchy of protein recruitment to chromatin.

Overall, my findings demonstrate that by using a template that mimics the native form of chromatin, which is not accessible by the widely used reductionist approaches, chromatin modifications can be analyzed in multiple dimensions. In combination with the new methods established here, my work will help to improve our understanding of epigenetic regulation of chromatin-associated processes.

1 Introduction

Epigenetic is the scientific field, which investigates heritable changes in gene expression that do not result from changes of the underlying DNA sequence, in other words an alteration of the phenotype by retaining the genotype. The term was first introduced by C. Waddington in the 1950s (Waddington, 1953) and referred in general to molecular processes affecting gene activity resulting in a particular phenotype. Over time, the definition of epigenetics narrowed and was defined as “An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence.” in 2009 (Berger et al., 2009). Those changes of chromatin can occur at several levels. Currently considered in literature are DNA methylation, covalent posttranslational histone modifications, histone variants and non-coding RNA associated silencing. All these processes are thought to be involved in the establishment and maintenance of certain chromatin states that alter chromatin structure and consequently gene regulatory processes.

1.1 Chromatin

1.1.1 Chromatin organization

The eukaryotic genome is located in the cell nucleus in form of a highly compacted structure called chromatin, made of DNA and its associated molecules including proteins and RNA. In the cell, chromatin exists at several levels of compaction, each of them regulating the accessibility of the DNA template to the transcription machinery and gene expression factors.

The first level of chromatin compaction consists in the linear arrangement of repeating uniform units of DNA and histone octamers. Histone octamers consist of two copies of each of the four core histone proteins H2A, H2B, H3 and H4, around which 147 base pairs (bp) of a left-handed DNA superhelix are wrapped in a 1.65 turn (Luger et al., 1997). These uniform units represent the basic packaging unit of chromatin and are referred to as nucleosome core particles (NCPs) (Kornberg, 1974; Olins and Olins, 1974). Histones are 11-15 kDa in size and very basic proteins. They belong to the most conserved proteins among eukaryotes. All core histones exhibit a common structural feature called the histone fold motif comprising three α -helices that are connected by two loops. These motifs mediate the protein-protein interactions between the histone pairs H2A-H2B and H3-H4 as well as the interaction between each histone pair and the DNA wrapped around the nucleosome. In presence of DNA or high salt concentration a tetramer of two copies of H3 and H4 and two H2A-H2B dimers self-assemble to a stable histone octamer (Luger et al., 1997). In contrast to the histone fold motif, the N- and C-terminal tails of histones

are far less structured. They form flexible and accessible domains protruding out of the nucleosome core particle and mediate multiple interactions with the neighboring nucleosomes as well as with non-histone proteins.

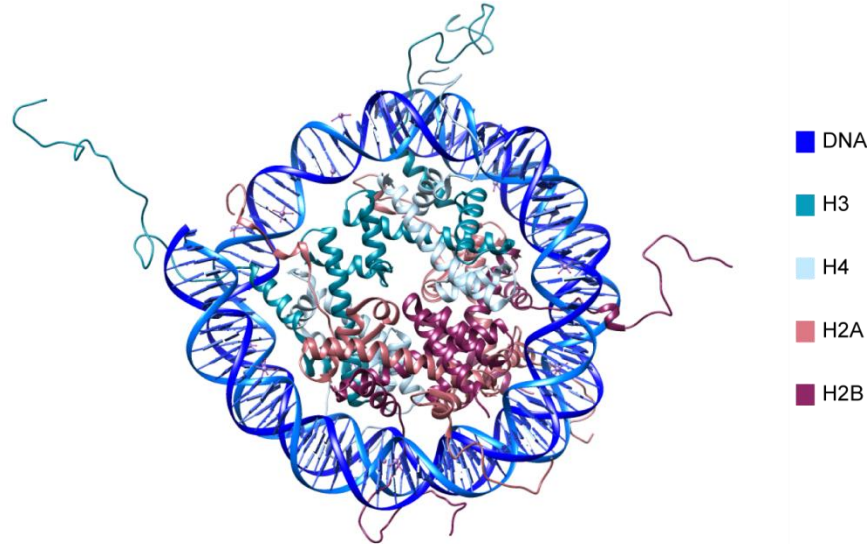


Figure 1.1 Structure of the nucleosome core particle.

Crystal structure at 1.9-Å (PDB_1KX5 (Davey et al., 2002)) visualized using Xlink Analyzer and Chimera (Kosinski et al., 2015; Pettersen et al., 2004). 147 bp of DNA (blue) is surrounding the histone octamer, which consists of two copies of each of the histones H2A (salmon), H2B (pink), H3 (turquoise) and H4 (light blue).

The next level of chromatin organization is stabilized by nucleosome to nucleosome arrangements. Nucleosomes are formed by both the nucleosome core particles and a linker DNA varying in length from 10 to 70 bp. Under physiological salt conditions and the presence of divalent cations an array of nucleosomes folds into a helical fiber of approximately 30 nm diameter. However, the detailed structure and the *in vivo* relevance of the 30 nm fiber is controversial and still under discussion (Kruithof et al., 2009; Schalch et al., 2005; Tremethick, 2007). Chromatin organization on levels beyond the 30 nm fiber is even less understood. Higher order compaction forming tertiary chromatin structures presumably involves inter-array interactions e.g. looping of chromatin fibers.

A fifth histone type, histone H1, contributes to stabilize the 30 nm fiber and higher order structures of chromatin by interacting with both, the linker DNA and the nucleosome core particles in a 1:1 ratio (NCP:H1). Nevertheless, it was shown that nucleosomal arrays fold into compact states even in absence of H1 (McBryant and Hansen, 2012; Woodcock et al., 2006).

Chromatin structure is not only influenced by the presence of H1. Variance in length and sequence of nucleosomal repeats, the presence of different histone variants and the pattern of posttranslational modifications (PTMs) of histones and DNA influence the structural formation of chromatin.

The structure of chromosomes can be generally distinguished in eu- and heterochromatin. The differentiation is ascribed to the heterogeneous staining intensities of chromosome areas observed with DNA staining dyes (Heitz, 1928). Heterochromatin represents chromosome regions formed by highly condensed chromatin and characterized by an intense DNA staining. These heterochromatic regions are thought to be distinctive for silenced gene loci (Dillon and Festenstein, 2002; Elgin and Grewal, 2003). Heterochromatin can be further distinguished into facultative and constitutive heterochromatin. Facultative heterochromatin mainly includes gene loci that are silenced in a time and tissue specific manner whereas constitutive heterochromatin is constituted by gene poor regions like repetitive satellite sequences, telomeres and centromeres. In contrast, euchromatin has a more relaxed structure. In consequence, staining by DNA specific dyes is less intense than the one observed for heterochromatin. The vast majority of transcriptionally active genes is present in euchromatin whose loose structure is thought to facilitate its accessibility for transcriptional factors. This explains the observations that constitutive heterochromatin remains condensed during the complete cell cycle while euchromatin undergoes decondensation during interphase.

The structural organization of chromatin is much more complex than the historical staining initially suggested and exists in many structural subtypes, which appeared to play key roles in the regulation of essential genomic functions like transcription, replication or DNA repair.

1.1.2 Chromatin modifications

PTMs of histones are one of the key mechanisms regulating chromatin structure and function. PTMs that occur on histones are biochemically diverse and include for instance serine/threonine/tyrosine phosphorylation, lysine acetylation, lysine and arginine methylation, lysine ubiquitination, lysine sumoylation, ADP-ribosylation, arginine citrullination and proline isomerization (Rothbart and Strahl, 2014). More than 150 different histone modification sites have been identified so far. The vast majority of modification sites are found within the N-terminal tails of histones although the number and relevance of posttranslational modifications at the histone globular domains are of increasing interest (Lawrence et al., 2016). Not only the different positions of histone modifications compose the degree of diversity, but also the level of the modification state. Methylation can occur in three degrees for lysine (mono-, di- and tri-) and arginine (mono-, symmetric and asymmetric di-) residues.

A model of how histone PTMs function in regulating chromatin structure was postulated with the histone code (Jenuwein and Allis, 2001; Strahl and Allis, 2000). This concept is based on antagonistic acting enzymes such as kinases/phosphatases, histone acetyltransferases (HATs)/histone deacetylases (HDACs) and lysine methyltransferases (KMTs)/lysine

demethylases (KDMs) that establish specific histone modification patterns, which are recognized and interpreted by non-histone proteins called histone modification readers.

The large variety of histone modifications provides a high potential for the regulation of chromatin organization and function. Histone modifications have been linked to cellular processes involved in transcriptional regulation, cell cycle process, DNA replication, DNA repair, splicing, epigenetic silencing and others (Izzo and Schneider, 2010; Kouzarides, 2007). Chromatin organization can be directly modulated by histone modifications that affect the interactions between nucleosomes. For instance, the incorporation of histone H4 lysine 16 acetylation (H4K16ac) was shown to reduce chromatin compaction and increase transcription (Akhtar and Becker, 2000; Shogren-Knaak et al., 2006) whereas H4K20 di- and tri-methylation have been shown to enhance chromatin condensation *in vitro* (Lu et al., 2008). In addition to these direct effects, histone modifications can act indirectly on chromatin organization by promoting or preventing the recruitment of chromatin binding proteins. Basically all DNA processes can be influenced by the regulation of chromatin access of chromatin modifying protein complexes or single factors, transcription factors and effector proteins that activate downstream signaling (Margueron et al., 2005; Torres and Fujimori, 2015; Wysocka et al., 2006).

Genomic distribution of histone modifications

The local enrichment of certain histone modification patterns is known to correlate with structural regions along the genome. Some histone modifications are highly enriched in heterochromatin and depleted from euchromatin while others are specific for euchromatin. Moreover, histone modification patterns correlate also with functional genomic elements along these chromatin subtypes. This observation allows using them as markers to differentiate certain genomic regions.

For example, in euchromatin histone marks like tri-methylation of histone H3 at lysine 4 (H3K4me3) and H3K27 acetylation are highly enriched at transcriptional start sites (TSS) of active genes (Kimura, 2013) while H3K36me3 and H3K4me1 are distributed within the gene body (Kimura, 2013). In contrast, heterochromatic inactive gene loci are mainly characterized by high levels of H3K9me3 and H3K27me3 as well as the presence of H3K9me2 and H3K27me2 (Kimura, 2013; Wang et al., 2009). Thus, it appears that inactive genes are depleted from active histone marks such as H3K4 methylation while in turn active gene loci appear to be mainly free of histone modifications specific for silenced genes. Nevertheless, those general rules can deviate depending on cell type and/or developmental stage. For example, in undifferentiated stem cells, bivalent chromatin domains were found to be simultaneously marked by both

H3K4me3 and H3K27me3, characteristic histone marks for eu- and heterochromatin, respectively (Bernstein et al., 2006; Vastenhouw and Schier, 2012). Those domains with a dual chromatin identity might have a potential function in the fine regulation of gene expression during embryonic development by the capacity of switching between a H3K4me3 transcriptionally active state and a H3K27me3 silenced gene state.

The development of techniques such as ChIP-on-chip and ChIP-Seq have considerably contributed to deepen our understanding of the global chromatin landscape by providing genome wide pictures of the histone mark distribution (Barski et al., 2007; Mikkelsen et al., 2007; Wang et al., 2008).

It was observed that histone lysine acetylation is generally found at transcribed regions of active genes, which links this modification to active transcription and gene expression in a more general way (Cui and Shi, 2016; Ucar et al., 2011; Wang et al., 2008). A set of distinct histone modifications is also generally linked to heterochromatin. H3K9me3, H4K20me3 and H3K27me3 have been characterized as predominant constitutive heterochromatin marks (Bannister and Kouzarides, 2011; Mikkelsen et al., 2007). In agreement with previous reports, which associated H3K9me3 and H4K20me3 with silencing of centromeres, transposons and tandem repeats, genome wide mapping revealed a strong enrichment of these marks at telomeric, satellite and long terminal repeats (Mikkelsen et al., 2007). The group of identified repressive marks included next to the tri-methylation sites of H3K9, H4K20 and H3K27 H3K27me2 and H3K9me2 while most other modifications correlated with transcriptional activation (Barski et al., 2007; Cui and Shi, 2016; Wang et al., 2008). Within heterochromatin not only constitutive regions are marked by histone modifications. H3K27me3 is the hallmark of facultative heterochromatin that includes regulatory chromatin elements like boundary elements and insulators (Bannister and Kouzarides, 2011; Van Bortle et al., 2012) but has also been detected at inactive enhancer elements. Surprising and in strong contrast to their di- and tri-methylation counterparts, the mono-methylated states of H3K9, H3K27 and H4K20 were found to be located mainly at active promoter sites and within the gene body, suggesting an association with transcriptional activity. Nevertheless, all three mono-methylation states have been already associated in the context of transcriptional repression. H3K27me1 was even shown to be significantly present in heterochromatin (Jacob et al., 2010).

Another histone modification assumed to be associated with heterochromatin is the symmetrical di-methylation of arginine 3 of histone H4 (H4R3me2). Genome wide analysis showed no prediction for either, active or silenced promoters (Barski et al., 2007). However, several studies associated H4R3me2 with transcriptional repression (Cui and Shi, 2016; Hou et al., 2008; Litt et al., 2009; Xu et al., 2010; Zhao et al., 2009).

Table 1.1 Histone modifications.

The table represents selected N-terminal histone modifications and their associated functions. The information of the table is based on (Izzo and Schneider, 2010; Lawrence et al., 2016).

histone	modification	function
H2A	H2AK4/5ac	transcriptional activation
	H2AK7ac	transcriptional activation
	H2AS1P	mitosis; chromatin assembly
	H2AK119P	spermatogenesis
	H2AK119uq	transcriptional repression
H2B	H2BK5ac	transcriptional activation
	H2BK11/12ac	transcriptional activation
	H2BK15/16ac	transcriptional activation
	H2BS14P	apoptosis
	H2BK120uq	spermatogenesis/meiosis
	H2BK123uq	transcriptional activation
H3	H3K4me2	permissive euchromatin
	H3K4me3	transcriptional elongation; active euchromatin
	H3K9me1	transcriptional activation/repression
	H3K9me2	transcriptional repression
	H3K9me3	transcriptional repression; imprinting; DNA methylation
	H3K27me1	transcriptional activation/repression
	H3K27me2	transcriptional activation/repression
	H3K27me3	transcriptional silencing; X-inactivation; bivalent genes/gene poising
	H3K36me3	transcriptional elongation
	H3R17me	transcriptional activation
	H3K4ac	transcriptional activation
	H3K9ac	histone deposition; transcriptional activation
	H3K14ac	transcriptional activation; DNA repair
	H3K23ac	transcriptional activation; DNA repair
	H3K27ac	transcriptional activation
	H3S10P	mitosis; meiosis; transcriptional activation
	H3t11/S28P	mitosis
H4	H4K20me1	transcriptional activation/silencing
	H4K20me3	heterochromatin
	H4R3me	transcriptional activation
	H4R3me2s	transcriptional silencing
	H4K5ac	histone deposition; transcriptional activation; DNA repair
	H4K8ac	transcriptional activation; DNA repair; transcriptional elongation
	H4K12ac	histone deposition; telomeric silencing; transcriptional activation; DNA repair
	H4K16ac	transcriptional activation; DNA repair

DNA methylation

In addition to posttranslational histone modifications chromatin can also be chemically modified on the DNA. In eukaryotes, DNA is predominantly methylated at carbon-5 of cytosine (5mC) in the context of CpG dinucleotides. CpG dinucleotides are found throughout the genome but are mainly located within CpG islands, which are characterized as short CpG rich regions that are often associated with TSSs of promoters, gene bodies and intergenic regions (Jones, 2012). CpG islands are usually unmethylated and associated with the euchromatic histone modification H3K4me3 (Guenther et al., 2007; Illingworth and Bird, 2009; Mikkelsen et al., 2007). However, most of the genome is CpG-deficient and essentially methylated.

In case DNA methylation of CpG islands occurs it is generally associated with transcriptional repression and connected with a functional role in imprinting, X-chromosome inactivation and transposon repression (Breiling and Lyko, 2015). Recent studies, however, suggest that DNA methylation can also be associated with promotion of transcription, indicating a far more complex functional diversity (Hu et al., 2013; Jin et al., 2012; Wu et al., 2010).

The establishment of DNA methylation is performed by three DNA methyltransferases comprising DNMT1, which has a strong preference for hemimethylated DNA while the two others, DNMT3A and DNMT3B, do not show such a preference and are therefore known as the *de novo* methyltransferases (Goll and Bestor, 2005). DNMT3A and DNMT3B establish DNA methylation in early development and together with DNMT1 participate in maintaining DNA methylation (Li et al., 2015; Liao et al., 2015).

1.1.3 Readout of chemical chromatin modifications

The dynamic control of chemical modifications of DNA and histones include the action of antagonistic working enzymes referred to as “writers” and “erasers”. Those enzymes are targeted to chromatin by “reader” domains that recognize specifically modified (or unmodified) amino acid residues and DNA. “Reader” domains can be found within the protein structure of “writers” and “erasers” or as part of an associated factor.

The most abundant binding domains involved in histone modification readout include bromodomains, Bromo-adjacent homology (BAH) domains, plant homeodomain (PHD) fingers, WD40 repeat domains, the “royal family” modules and 14-3-3 domains.

Bromodomains are acetyl-binding domains that specifically recognize ϵ -N-acetylated lysine residues (Owen et al., 2000). They are present in a variety of chromatin-associated factors, including HAT and HMT enzymes, ATP-dependent helicases and also transcription factors (Ferri

et al., 2016). Bromodomains often occur in tandem with other domains within the same protein. They are most frequently found associated together with PHD fingers, but also with PWWP, SET and BAH domains (Sanchez et al., 2014).

BAH domains have been found to act as protein–protein interaction modules, histone lysine methylation recognition and nucleosome binding modules. There are present in many proteins associated with DNA and histone modifications, such as DNMT1, MTA1 and the origin recognition complex 1 (Orc1) (Yang and Xu, 2013). For instance, the BAH domain was shown to be important in mediating the preferential binding of OCR1 to H4K20me2 (Kuo et al., 2012). BAH domains are essentially found in factors connected to processes involved in DNA methylation, replication and transcriptional regulation. Therefore it is also not surprising to find them often together with reader modules known to bind histones and DNA, such as Bromo, SANT and PHD domains.

PHD domains are structurally conserved modules that act as readers of methylated lysine on histone tails as well as non-histone proteins (Musselman and Kutateladze, 2011). In rare cases, PHD domains have been shown to interact with methylated arginine and acetylated lysine. For instance, the PHD finger of RAG2 preferentially interact with symmetrically methylated H3R2 whereas the tandem PHD finger of DPF3b was shown to bind to H3K14ac (Ramon-Maiques et al., 2007; Zeng et al., 2010). As most of the other reader modules, PHD fingers are flanked by additional reader domains, which appear to be important in combinatorial readout of histone modifications.

The “royal family” modules are a group of structurally related protein motives, including the Tudor, PWWP, chromatin-binding (Chromo) and malignant brain tumor (MBT) domains that recognize lysine methylation by aromatic cage pockets (Maurer-Stroh et al., 2003). Besides individual exceptions of the PHD and the WD40 domain, Tudor domains are the only arginine binding motives known so far (Gayatri and Bedford, 2014). Some of the proteins harboring a Tudor domain have been shown to interact with specific histone methylation sites as well. While the Tudor domains of PHF1 and PHF19 interact specifically with H3K36me3, the Tudor domains of JMJD2A and UHRF1 can target several modification states, alone or in combination with a second binding motive, respectively (Lu and Wang, 2013).

PWWP domains contain a highly conserved Pro-Trp-Trp-Pro motif and were initially identified as non-specific DNA-binding domains. However, recently the PWWP domain of Brf1 has been demonstrated to specifically associate with H3K36me3 (Vezzoli et al., 2010).

Chromodomains are very well characterized as they were discovered among the first histone binding folds. Although chromodomains are structurally similar, they can be divided into different

subclasses comprising HP1/CBX chromodomains, chromobarrel, chromo shadow and chromo-ATPase/helicase-DNA-binding (CHD) domains (Yap and Zhou, 2010). The proteins heterochromatin protein 1 (HP1) and Polycomb were the first factors characterized as specifically recognizing tri-methylated lysine 9 of H3 and tri-methylated lysine 27 of H3 via their chromodomain (Patel, 2016; Yap and Zhou, 2010). CBX chromodomains show a similar preference for those two modification marks although they have been additionally functionally associated with H3K27me2. Interestingly, the chromo shadow domains do not interact with methylated histones but were shown to act as protein-protein interaction modules (Brasher et al., 2000; Cowieson et al., 2000) while the MSL3 chromobarrel domain was shown to bind to mono-methylation of H4K20 only in the presence of DNA (Kim et al., 2010).

Proteins harboring MBT repeats are functionally involved in processes like mitosis, tumor suppression and preferentially bind to mono- and di-methylated lysines of histones (Bonasio et al., 2010). The MBT modules as well as the other members of the “royal family” can occur and act as tandem reader modules (Patel, 2016).

WDR 40 motifs are another family of modules acting in form of tandem repeats, which bind to methylated as well as unmethylated histones. One of the best studied representatives is the protein WDR5 (Patel, 2016; Yap and Zhou, 2010).

Histone phosphorylation mainly occurs at serine residues that can be recognized by 14-3-3 domains as well as the tandem breast cancer susceptibility (BRCT) domain. For instance, the BRCT domain of the protein Breast cancer associated 1 (BRCA1) has been shown to be important for its function in DNA repair (Glover, 2006; Yap and Zhou, 2010).

5mC methylated DNA is also recognized by certain protein-binding motifs. So far, three different binding motifs were identified – the meCpG-binding domain (MBD), SET and RING-associated (SRA) domains and Kaiso and Zfp57 zinc finger proteins (Patel, 2016). Kaiso can recognize symmetrically modified 5mC in the context of tandem CpG dinucleotides via a zinc finger motif (Prokhortchouk et al., 2001). In mammals, SRA domains are restricted to the UHRF1 protein family comprising UHRF1 and UHRF2 and have been demonstrated to bind hemimethylated 5mCpG/CpG sites (Bronner et al., 2007; Unoki et al., 2004). MBD domains were shown to recognize symmetrically methylated CpG dinucleotides (Nan et al., 1993; Ohki et al., 2001). Independent of the DNA methyl binding domain, 5mC-binding proteins have been shown to recruit proteins associated with transcriptional repression (Patel, 2016).

1.1.4 Multivalent readout and crosstalk of chemical chromatin modifications

The diversity of recognition of chemical chromatin modifications is already enormous when only considering the range of binding folds identified and their specificity, which is supported by their arrangement within a polypeptide and the flanking amino acid sequences of the PTMs to be recognized. The diversity strongly increases considering additional processes of multivalent readout. The discovery of the linkage of multiple binding domains, either within a single protein or multi protein complexes, strongly supports the assumption of complex readout. In particular, the readout of a certain modification can be influenced by adjacent PTMs offering countless PTM combinations functioning in a synergistically as well as antagonistically manner.

Several mechanisms of combinatorial readout of chromatin modifications are conceivable. The combinatorial readout can either take place on the same histone tail or on different histone tails, referred as *in cis* and *in trans*, respectively. Readout *in trans* happens either on a particular histone or at the nucleosomal level. The latter can be further distinguished between intra- and inter-nucleosomal combinatorial readout. Additionally, multivalence can also occur concomitant DNA and histone recognition. Binding modules involved in combinatorial readout can be coupled by protein-protein interactions in multiple protein complexes or directly within a single protein. Possibilities of protein-binding modes reflecting the introduced mechanisms are illustrated in figure 1.2.

To date, the combinatorial readout *in cis* and *in trans* by paired protein-binding modules within one protein could be demonstrated for several proteins. For instant, combinatorial readout *in cis* via a PHD-Bromo cassette has been shown for TRIM24 and TRIM33, which bind unmodified H3 in parallel with H3K23ac and H3K9me3 concomitantly with H3K18ac, respectively (Tsai et al., 2010; Xi et al., 2011). The tandem Tudor-PHD finger cassette of UHRF1 has also been shown to bind the N-terminus of H3 *in cis* (Arita et al., 2012). Combinatorial binding *in trans* is far less studied. One of the best studied examples is the intra-nucleosomal binding of the PHD-Bromo cassette of BPTF that associates with H3K4me3 and H4K16ac (Ruthenburg et al., 2011).

Besides paired protein-binding modules within one protein, crosstalk can be addressed on the level of indirect protein recruitment by protein-protein interactions. Chromatin modifying protein complexes can contain multiple “readers” providing a wide range of binding motifs in combination with “writers” and “erasers”. The local accumulation of such complexes provides the advantage of a higher degree of specificity and the ability to establish a complex series of modification patterns. The HBO1 complex for instance contains several PHD finger modules that act cooperatively to regulate its acetyltransferase activity (Musselman et al., 2012; Torres and Fujimori, 2015). In a similar fashion, WD40 repeats, PHD and Tudor domains of the PRC2

complex regulate its methyltransferase activity. Different posttranslational histone modifications have been shown to specifically recruit those protein complexes and additionally provided first insights into the multivalent impact of DNA and histone modifications (Bartke et al., 2010; Bluhm et al., 2016; Engelen et al., 2015; Kunowska et al., 2015; Nikolov et al., 2011; Vermeulen et al., 2010).

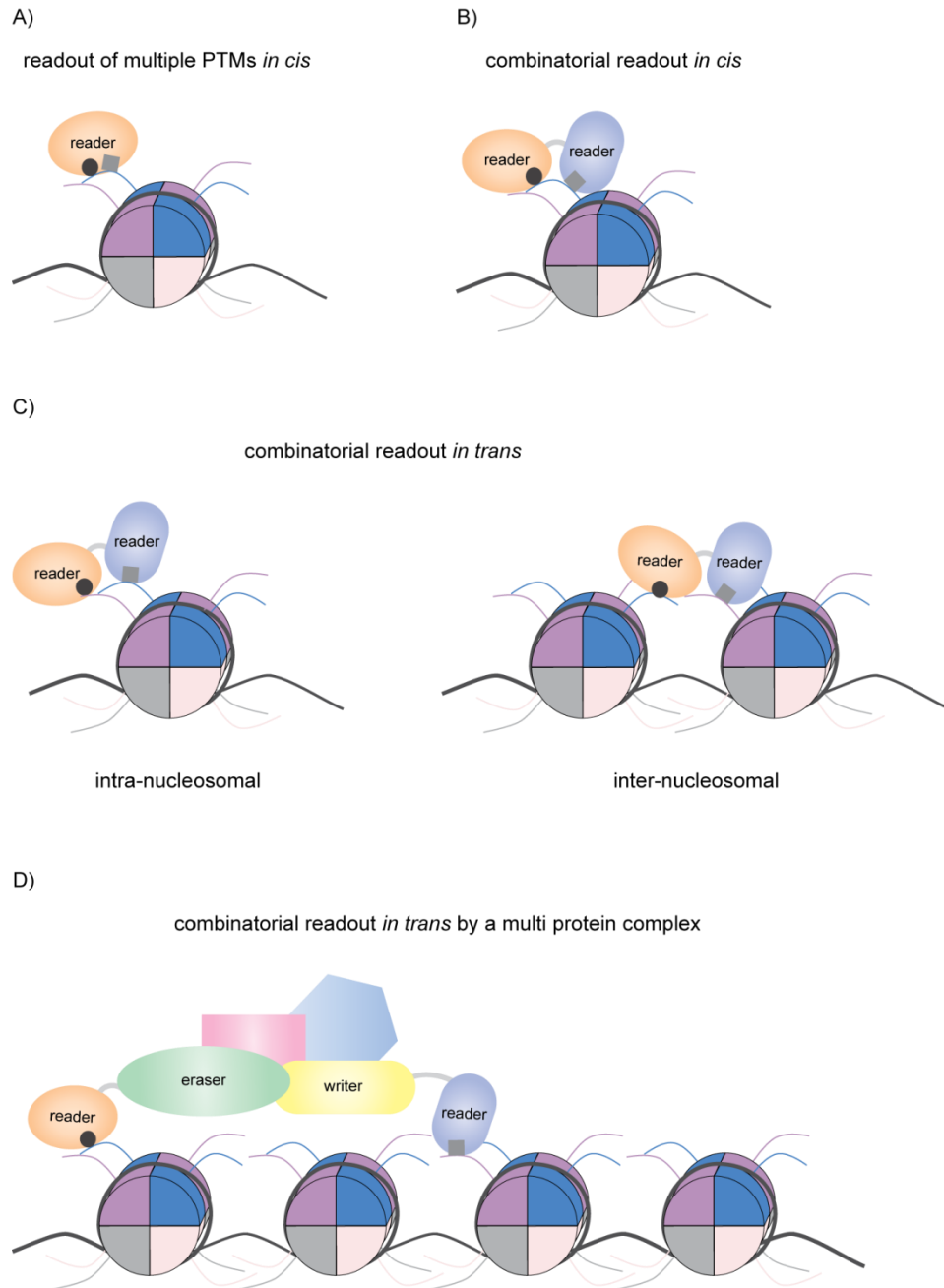


Figure 1.2 Combinatorial readout of posttranslational histone modifications.

A) Readout of two histone modifications on the same histone tail by a single protein. B) Combinatorial readout *in cis*. Histone PTMs on the same histone tail are recognized by two different proteins. C) Combinatorial readout *in trans*. The readout of two PTMs at different histone tails can take place either within one nucleosome (intra-nucleosomal) as shown in the left panel or on different nucleosomes (inter-

nucleosomal) as represented in the right panel. D) Combinatorial readout *in trans* facilitated by a multi protein complex. The figure is adapted from (Musselman et al., 2012).

Nevertheless, recent knowledge of PTM recognition mechanisms is mainly based on specific binding of single proteins to distinct histone modifications or combinations thereof. The extent of the influence various modification patterns have on protein binding to chromatin and the functional consequences thereof are still not fully understood. Toward understanding the epigenetic language of histone and DNA modifications, the assignment of functional consequences and therefore the identification of the complement of proteins regulated by individual chromatin modifications and combinations is of immense importance.

1.1.5 Chromatin arrays for systematic analysis of chemical modification readout

The question of how distinct histone modifications influence the protein recruitment to chromatin and more globally the chromatin interactome consists in the field of epigenetics since many years. To approach this question several strategies have been developed over years. Modified histone N-terminal peptides are one of the major tools used in approaches ranging from array- to affinity-based protein interaction studies (Kim et al., 2006; Wysocka, 2006). In general, affinity purifications strategies enable an unbiased, modification-dependent identification of chromatin-associated factors and macromolecular protein complexes. The experimental design is based on the incubation of unmodified or modified N-terminal histone peptides with nuclear extracts followed by the identification of the bound proteins by mass spectrometric analysis (Wysocka, 2006). Recently, this approach has seen important modification by coupling it with quantitative mass spectrometry (Oda et al., 2010; Vermeulen et al., 2010). However, the characterization of histone PTMs binding proteins has, so far, essentially relied on the use of *in vitro* synthesized modified histone N-terminal peptides as bait in affinity purification that may not reflect faithfully the diversity and complexity of the PTM-regulated histone interactome. One strategy to overcome this problem has consisted in using mononucleosomes or even oligonucleosomal arrays (Bartke et al., 2010; Nikolov et al., 2011). Nucleosomal arrays are of particular interest as they offer a far more physiologically relevant model compared to the peptide-based strategy. For example, chromatin arrays allow to (i) control histone PTM and underlying DNA sequences, (ii) to investigate the binding of proteins within the nucleosomal context including the identification of proteins binding only in presence of more than one nucleosome, (iii) characterize the crosstalk between histone modifications by inserting two or more PTMs on one single chromatin array, (iv) investigate the crosstalk between PTMs of histones and DNA.

The assembly of homogenous recombinant nucleosomal arrays has been well described in the literature (Dyer et al., 2004; Huynh et al., 2005; Luger et al., 1999). A general overview is shown

in figure 1.3. The system is based on histone octamers that are uniformly position on a defined DNA template (Huynh et al., 2005).

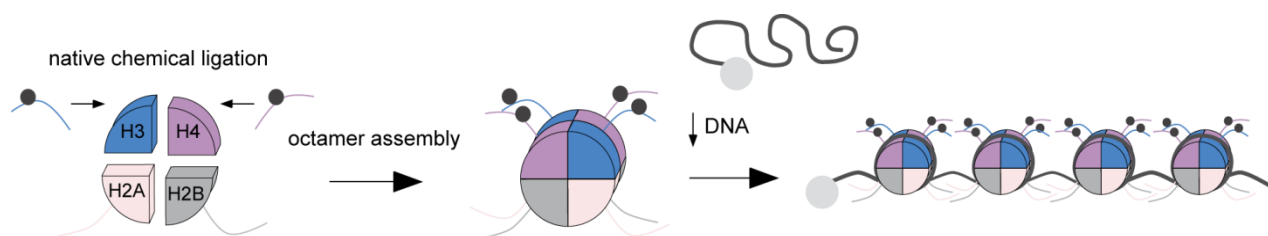


Figure 1.3 Design of recombinant nucleosomal arrays.

Recombinant full length histones, either unmodified or modified by native chemical ligation, were assembled to histone octamers via dialysis against high salt. Biotinylated 12x200x601 DNA templates and histone octamers containing modified histones of choice were mixed and reconstituted to 12mer chromatin arrays by dialysis over a salt gradient.

The application of chromatin based affinity purification of an unmodified and a chemically modified chromatin array in direct comparison demonstrated that nucleosomal arrays are well suited for the identification of specific protein-binding interactomes of chromatin arrays in the context of defined chemical modifications, when coupled with relative quantification by mass spectrometry (Nikolov et al., 2011). Thus, this method has the potential to elucidate the complex protein interactomes of diverse chromatin modification patterns and can provide insights into epigenetic regulated biological pathways and histone modification crosstalk.

1.2 Mass spectrometry

1.2.1 Mass spectrometry based protein identification of complex samples

Mass spectrometry is a powerful method especially in terms of global protein identification, which has allowed the characterization of organelle proteomes, signaling pathways and complete protein complexes. Although mass spectrometry has a long history in science, it is only during the last 20 years that it has been developed to the standard method for protein identification and quantification of complex samples in proteomics.

The principle of identification and quantification of molecules by mass spectrometry is based on the mass-to-charge ratio of the molecules. The molecule of interest is transferred into the gas phase and ionized. These ions can be separated according to their mass-to-charge ratio (m/z) by ion acceleration in an electric or magnetic field. Ion acceleration is later converted in ion current and gives the signal to be detected.

A standard proteomics workflow for MS based protein identification is carried out by tandem mass spectrometry comprising four main successive steps – (i) protein proteolysis, (ii) m/z detection of intact ions, (iii) precursor ion selection and fragmentation, (iv) peptide identification followed by protein identification using protein sequence databases (figure 1.4).

The proteomic workflow starts with the proteolysis of the proteins of interest. For the reason that peptides possess solubility in a wider range of solvents and their masses can be determined with high accuracy, the complex protein sample is proteolytic digested into peptides prior mass spectrometric analysis. Consequently, the degree of complexity increases even more. To ensure a comprehensive, unambiguous identification, including the detection of low-abundance species, a sufficient separation is required. While SDS-PAGE separation can be a first step to reduce the degree of complexity at the protein level, the application of liquid chromatography (LC) is successfully in use to reduce the complexity at the peptide level. The separation is achieved by nanoliter flow rate reversed-phase C18 chromatography coupled to direct elution into an electrospray ionization (ESI) mass spectrometer. Additional LC systems are commonly in use as well as separation based on capillary electrophoresis and the application of multidimensional separation, successfully increasing the number of identified peptides (Di Palma et al., 2012; Heemskerk et al., 2016; Yates, 2004).

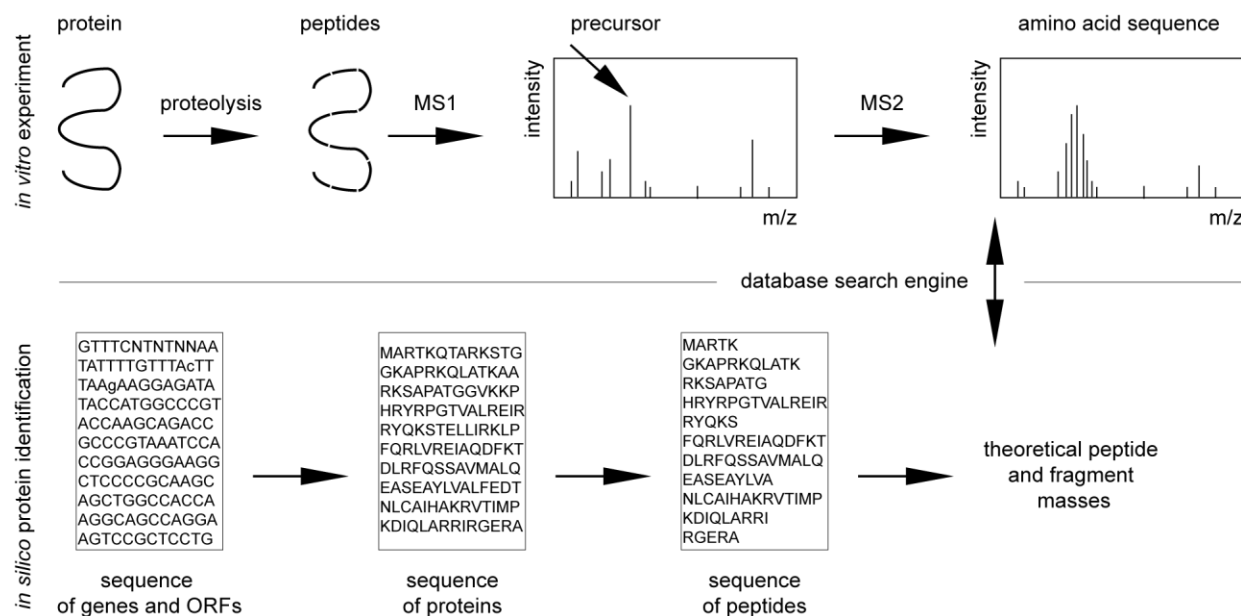


Figure 1.4 Workflow for protein identification by MS.

Proteins are hydrolyzed to peptides using a specific protease. The mass-to-charge ratios of the peptides are determined by mass spectrometry (MS1). Based on the first measurement precursor ions are selected and fragmented (MS2). For protein identification the information obtained from the MS1 and MS2 spectra are compared to the theoretical masses of a database by a search engine. The figure is adapted from (Schmidt, 2010).

The majority of tandem mass spectrometry experiments in proteomics use data-dependent-acquisition (DDA) for detection of m/z ratios. In a first step, the MS1 spectrum is generated by a complete scan over a distinct mass range that records m/z ratios of all ions eluting from the LC at a certain time point. The most abundant precursor ions are selected for fragmentation, commonly done by low-energy collision induced dissociation (CID). The scans of the fragmented ions provide the MS2 (or MS/MS) spectra. The DDA cycle acquires data over the length of the LC elution gradient.

The peptide identification is performed *in silico* by database search engines. A protein sequence database provides the theoretically mass of proteolytically digested peptides. The mass information resulting from the matched MS1 and MS2 spectra is used to search for peptide candidates in the database with corresponding theoretically masses. A certain mass tolerance setting according to the mass accuracy of the mass spectrometer is used and the results are reported according to statistical ranking.

1.2.2 Relative quantification by Stable Isotope Labeling by Amino acids in Cell culture (SILAC)

Quantitative mass spectrometry is of major importance in the field of MS based proteomics. Two general approaches have been developed over the last years – stable isotope-based and label-free methods. Label-free methods are based on comparison of identical peptides and proteins between multiple measurements and rely on either spectrum count or MS1 intensity based approaches (Bantscheff et al., 2012). With vast technical progress, label-free approaches have become more popular nowadays but still struggle with certain drawbacks. They require several replicates and rely on high reproducibility in terms of sample preparation and separation steps prior to mass spectrometric measurements. Stable isotope-based mass spectrometry relies on the nearly identical physicochemical properties of heavy stable isotopes and their corresponding light isotopes. Relative quantification is achieved by comparing the MS intensities of unlabeled peptides to their corresponding heavier isotopic labeled peptides. Consequently, the method allows the measurement of different labeled peptides (and proteins) within the same sample, thus reducing errors from sample preparation and ion suppression by mass spectrometers in different MS runs.

Isotope labels can be incorporated chemically, enzymatically or metabolically (Ong and Mann, 2005). One of the most popular metabolic labeling approaches is Stable Isotope Labeling by Amino acids in cell culture (SILAC) (Ong et al., 2002). The incorporation of heavy isotopes takes place during cellular or organismal growth with the consequence of a certain limitation in the range of applicable samples. This drawback was reduced with the recent development of super SILAC and spike-in SILAC allowing the application of SILAC to tissues and body fluids (Geiger

et al., 2010; Geiger et al., 2011). Another disadvantage is the number of samples that can be compared using SILAC. While a few years ago the complexity of analysis caused restrictions the actual limitations are reasoned in the limited number of parallel labels for heavy labeled amino acids. Widely in application are the stable isotope labeled amino acids lysine and arginine, containing C13 and N15. Using labeled lysines and arginines up to 5-plex SILAC experiments can be performed (Ong et al., 2002). Nevertheless, SILAC has proven to be a method of choice as it has been proven to be highly accurate and more robust compared to other stable isotope-based methods and therefore suitable for affinity purification strategies (Chen et al., 2015).

Quantitative MS-based proteomics as large scale MS-based approaches in general require specific algorithms for peptide/protein identification and quantification. MaxQuant is a quantitative software packages that enables protein identification and quantification in the context of SILAC based approaches but also supports quantification of additional label-based and label-free approaches (Cox and Mann, 2008; Cox et al., 2011). The software incorporates all steps of data analysis, starting from processing MS spectra to final statistical evaluation. The analysis pipeline starts with high intensity-weighted estimation of masses for peptide peaks (assignment of 3D peaks based on intensity, m/z and elution time), the determination of reliably isotope patterns and the assignment of SILAC peptide pairs. After improvement of mass accuracy, peptide and protein identification is done using an integrated database search engine. The last steps are directed at protein quantification based on the enrichment ratios of individual peptide pairs. The software package Perseus enables statistical evaluation and visualization.

1.2.3 Protein-protein cross-linking mass spectrometry

Cross-linking mass spectrometry (XL-MS) is a powerful approach, which allows investigating structural as well as physical interaction sites of proteins. A bifunctional cross-linker is used to covalently link proteins in physical close proximity engaged in non-covalent interactions. Commonly, cross-linkers used in MS possess two cross-linking reactive groups that are connected by a spacer but can have properties like variable spacer length, cleavability, composition, solubility, isotope labeling and photo reactivity (Paramelle et al., 2013). Another very important property is the specificity of a cross-linker that ranges from low to high specificity and can be modulated by altering the cross-linking conditions such as pH and the protein-cross-linker ratio. One of the most popular cross-linking reagents in use is bis(sulfosuccinimidyl) suberate (BS3). BS3 is a chemical cross-linker with a spacer arm length of 11.4 Å and belongs to the highly reactive N-hydroxysuccinimide (NHS) ester that target nucleophiles. It is a homobifunctional cross-linker that can react with primary amines in the side chains of lysine (K)

residues and the N-terminus of polypeptides resulting in stable amide bonds (Sinz, 2006; Tran et al., 2016).

After cross-linking, non-cross-linked peptides are still the overrepresented peptide species in a given sample (Sinz, 2006). Thus, compared to linear peptides, cross-linked peptide are less abundant and therefore less likely to be identified by mass spectrometric measurements. To overcome these issues, different enrichment strategies for cross-linking containing species can be applied on the protein as well as on the peptide level. Based on the property of a higher molecular weight of cross-linked species separation according to size can successfully enrich for both, cross-linked proteins and cross-linked peptide pairs. Depending on experimental design and sample preparation SDS-PAGE and size exclusion chromatography are applicable methods on the protein level. Size exclusion chromatography also has been shown to be applicable on the peptide level (Leitner et al., 2012). Alternatively, separation of cross-linked peptides was successfully achieved applying strong cation exchange chromatography as they exhibit more charges compared to linear peptides (Fritzsche et al., 2012).

Still, one of the major challenges in XL-MS is the identification of cross-linked peptides and proteins. Database search engines used for identification of proteins based on linear peptides cannot be applied for the analysis of cross-link spectra. The main reason is due to the fact that both, MS1 and MS2 spectra do not correspond to one peptide but to all possible combinations of cross-linked peptide pairs. The consequence is a dramatic increase of the database used for the search associated with increased risk of random assignments and higher false discovery rates. To date, two data processing tools, xQuest/xProphet and pLink, have been developed that search for cross-links against large databases exceeding a number of 30 to 40 proteins, implementing a false discovery rate control (Purcell et al., 2007; Rinner et al., 2008; Walzthoeni et al., 2012). Both have been proven to be suitable tools, sensitive enough to study protein-structure and protein-protein-interactions. The disadvantage of xQuest/xProphet is the necessity of labeled cross-linker, thus pLink was the software used for my studies.

1.3 Hypothesis and objectives of the presented thesis

Chromatin structure adopts two main configurations: a relaxed state containing most of the expressed genes and referred to as euchromatin and a condensed state associated with transcriptional silencing called heterochromatin. Both chromatin states exhibit histone posttranslational modifications found to be specifically associated with either eu- or heterochromatin. These chromatin modifications form a complex language whose understanding has been subject of constant efforts over the past decade. To date, the vast number of histone modification sites that have been identified, their distribution across the genome as well as their specific binders have provided considerable information on how chromatin functions can be regulated. However, a comprehensive overview of the proteomes associated with defined chromatin modifications and their functional outcomes were not achieved so far.

The hypothesis that has set the basis of my thesis work was that the biochemical status of chromatin domains is brought about by single chromatin modifications and combinations thereof that regulate the recruitment and the exclusion of specific networks of chromatin-associated proteins. One important question was whether a set of factors, common to all modifications associated with a given chromatin stage, would be responsible of the eu- or heterochromatin identities.

To address these questions, my primary objective was to characterize the proteomes associated with the majority of the known heterochromatic chromatin modifications and combinations thereof, by using a strategy based on chromatin affinity purification coupled to quantitative mass spectrometry. One of the first aims was to implement a statistical function that defines a fold enrichment cutoff meeting the requirements for the comparison of modification-specific regulated factors identified by individual performed experiments. Proteins regulated by specific modifications provided the basis for the identification of the biological functions of each of the ten heterochromatic modifications and the information regarding their functional relationships. The scope of my work has been broadening by introducing the question whether combinations of two chromatin modifications have an impact on the biochemical status of chromatin domains different from the individual modifications. To investigate this point, the combinations of H3K9me3 and H4K20me3 as well as H3K9me3 and CpG methylated DNA were investigated. Lastly, a new workflow coupling ChAP with cross-linking mass spectrometry was establish in order to map the physical interactions among chromatin-associated proteins with the aim to determine the hierarchy of protein recruitment to chromatin.

2 Materials and Methods

2.1 Material and reagents

2.1.1 Laboratory equipment

ÄKTA Explorer/Purifier/micro	GE Healthcare, Buckinghamshire (UK)
Balances	Metler-Toledo, Gießen (DE)
BBD 6220 CO ₂ incubator	Heraeus, Hanau (DE)
Bioreactor 5L with ez-Control	Applikon, Schiedam (NL)
Centrifuge Cryofuge 6000i	Heraeus, Hanau (DE)
Centrifuge Sorvall Evolution RC	Thermo Scientific, Braunschweig (DE)
Centrifuges tabletop 5415R/5810R	Eppendorf, Hamburg (DE)
ChemiDoc MP System	Bio-Rad, München (DE)
Electrophoresis power supplies	Bio-Rad, München (DE)
HP1100 and HP1200 LC systems	Agilent, Santa Clara (USA)
Laminar flow clean bench	Heraeus, Nahau (DE)
LTQ-Orbitrap Velos	Thermo Fischer Scientific, Bremen (DE)
LTQ-XL	Thermo Fischer Scientific, Bremen (DE)
Mini Trans-Blot system	Bio-Rad, München (DE)
Mini-PROTEAN Tetra PAGE cell	Bio-Rad, München (DE)
NanoDrop ND-1000	Peqlab, Erlangen (DE)
Peristaltic pump	Ismatec, Glattburgg (CH)
pH meter	Metler-Toledo, Gießen (DE)
Q-Exactive	Thermo Fischer Scientific, Bremen (DE)
Sonication bath SONOREX Super	BANDELIN Electronic, Berlin (DE)
Sorval SA600 rotor	Thermo Scientific, Braunschweig (DE)
Sorval SS34 rotor	Thermo Scientific, Braunschweig (DE)
SpeedVac Savant SPD121P	Thermo Scientific, Braunschweig (DE)
Sub-Cell-GT agarose gel electrophoresis	Bio-Rad, München (DE)
Thermocycler eppgradientS	Eppendorf, Hamburg (DE)
Thermomixer Comfort	Eppendorf, Hamburg (DE)
Water bath TW12	Julabo, Selbach (DE)
XCell Sure Lock Mini NuPAGE cell	Invitrogen, Karlsruhe (DE)

2.1.2 Consumables and plastic ware

Amicon Ultra centrifugal filter devices (MWCO 3 and 10 kDa)	Millipore, Billerica (USA)
Phase Lock Heavy Tubes (2-, 15-, 50 ml)	5PRIME, Hamburg (DE)
Slide-A-Lyzer dialysis units and cassettes (MWCO 3500, 7000, 10000)	Pierce/Thermo Scientific, Rockford (USA)
Spectra/Por dialysis membrane (MWCO 3500 and 10000)	Spectrum Laboratories, Rancho Domingues (USA)

2.1.3 Chemicals

(2-bromoethyl)-tri-methylammonium bromide	Sigma-Aldrich, Steinheim (DE)
(2-chloroethyl)-dimethylammonium chloride	Merck, Darmstadt (DE)
(2-chloroethyl)-methylammonium chloride	Karl Industries, Aurora (USA)
2-mercaptoethanol	Sigma-Aldrich, Steinheim (DE)
4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	VWR, Poole (DE)
4-Mercaptophenylacetic acid (MPAA)	Sigma-Aldrich, Steinheim (DE)
Acetic acid	Merck, Darmstadt (DE)
Acetonitrile, LiChrosolv	Merck, Darmstadt (DE)
Acrylamide/Bisacrylamide (37.5:1)	Merck, Darmstadt (DE)
Agarose	Serva, Heidelberg, (DE)
Ammonium hydrogen carbonate	Fluka, Buchs (CH)
Ammonium peroxodisulfate	AppliChem, Darmstadt (DE)
Bis(sulfosuccinimidyl) suberate (BS3)	Thermo Scientific, Schwerte (DE)
Boric acid	Merck, Darmstadt (DE)
Bovine serum albumin (BSA)	Sigma-Aldrich, Steinheim (DE)
Bromophenol blue	Serva, Heidelberg (DE)
Cleland's reagent (DTT, for MS analysis)	Calbiochem, Darmstadt (DE)
Coomassie Brilliant Blue G-250	Fluka, Buchs (CH)
D/L-Methionine	Sigma-Aldrich (DE)
Deoxynucleotide-5'-phosphate (dATP, dCTP, dGTP, dTTP)	Roth, Karlsruhe (DE)
Dipotassium hydrogen phosphate (K_2HPO_4)	Roth, Karlsruhe (DE)
Dithioerythrol (DTE)	Roth, Karlsruhe (DE)
Dithiothreitol (DTT)	Alexis Biochemicals, Farmingdale (USA)

Ethanol	Merck, Darmstadt (DE)
Ethidium bromide	Roth, Karlsruhe (DE)
Ethylendiamine tetraacetate (EDTA)	Roth, Karlsruhe (DE)
Formic acid (FA)	Fluka, Buchs (CH)
Glycerol	Merck, Darmstadt (DE)
Guanidine hydrochloride	Sigma-Aldrich, Steinheim (DE)
Guanidine hydrochloride ($\geq 99.5\%$)	Roth, Karlsruhe (DE)
Hydrochloric acid (37% HCl)	Merck, Darmstadt (DE)
Iodacetamide (IAA)	Sigma-Aldrich, Steinheim (DE)
LB Broth	MOBIO, Hamburg (DE)
Magnesium chloride (MgCl_2)	Merck, Darmstadt (DE)
Methanol, LiChrosolv	Merck, Darmstadt (DE)
N,N,N',N'-Tetramethylethylenediamid (TEMED)	Sigma-Aldrich, Steinheim (CH)
Non-fat dry milk powder	Regilait, Saint-Martin-Belle-Roche (FR)
Ortho-Phosphoric acid	Merck, Darmstadt (DE)
Phenol:Chlorophorm:Isoamil alcohol (PCI) [25:24:1]	Roth, Karlsruhe (DE)
Phenyl-methylsulfonyl fluoride (PMSF)	Roche, Mannheim (DE)
Polyethylene glycol 6000 (PEG-6000)	Merck, Darmstadt (DE)
Ponceau S	Sigma-Aldrich, Steinheim (DE)
Potassium chloride (KCl)	Merck, Darmstadt (DE)
Potassium dihydrogen phosphate (KH_2PO_4)	Roth, Karlsruhe (DE)
RapiGest	Waters (USA)
S-(5'-adenosyl)-L-methionine (SAM)	New England Biolabs, Ipswich (USA)
Sodium acetate	Roth, Karlsruhe (DE)
Sodium azide (NaN_3)	Alfa Aesar, Massachusetts (USA)
Sodium chloride (NaCl)	Merck, Darmstadt (DE)
Sodium dodecyl sulfate (SDS)	VWR, Poole (DE)
Sodium hydrogen phosphate (Na_2HPO_4)	Merck, Darmstadt (DE)
Sodium hydroxide (NaOH)	Merck, Darmstadt (DE)
Tris(hydroxymethyl)amino ethane (Tris base)	Roth, Karlsruhe (DE)
Triton X-100	Merck, Darmstadt (DE)
Tween-20	Sigma-Aldrich, Steinheim (DE)
Urea	Merck, Darmstadt (DE)
Water, LiChrosolv	Merck, Darmstadt (DE)

2.1.4 Chromatographic and affinity material

C18 reversed phase precolumn (0.15 mm ID x 20 mm, Reprosil-Pur120 C18-AQ 5 µm)	Dr. Maisch, Ammerbuch-Entringen (DE)
HiLoad Superdex 200 10/300	GE Healthcare, Buckinghamshire (UK)
HiLoad Superdex 200 16/60	GE Healthcare, Buckinghamshire (UK)
Superdex 200 Increase 3.2/300	GE Healthcare, Buckinghamshire (UK)
Superdex Peptide 3.2/300	GE Healthcare, Buckinghamshire (UK)
HiTrap SP HP 1ml	GE Healthcare, Buckinghamshire (UK)
PD-10 columns	GE Healthcare, Buckinghamshire (UK)
Q-Sepharose XK26/20	GE Healthcare, Buckinghamshire (UK)
Reposil AQ-3/5µm / 300Å	Dr. Maisch, Ammerbuch (DE)
SilicaTip emitters	New Objective, Woburn (USA)
SP-Sepharose XK26/20	GE Healthcare, Buckinghamshire (UK)
Streptavidin MagneSphere particles	Promega, Mannheim (DE)

2.1.5 Cell culture media and materials

DMEM High Glucose ([w/o] Lys, [w/o] Arg)	PAA Laboratories, Colbe (DE)
Fetal bovine serum, dialyzed	PAA Laboratories, Colbe (DE)
L-Arginine (Arg0)	Sigma-Aldrich, Steinheim (DE)
L-Arginine, ¹³ C ₆ (Arg6)	Euriso-top, Saarbrücken (DE)
L-Lysine (Lys0)	Sigma-Aldrich, Steinheim (DE)
L-Lysine, ² D ₄ (Lys4)	Euriso-top, Saarbrücken (DE)
Penicillin/Streptomycin 100 x	PAA Laboratories, Colbe (DE)

2.1.6 Commercial kits

1 Kb Plus DNA Ladder	Invitrogen, Karlsruhe (DE)
ECL Pus Western Blotting Detection System	GE Healthcare, Buckinghamshire (UK)
ECL Western Blotting Detection System	GE Healthcare, Buckinghamshire (UK)
Hybond ECL nitrocellulose membrane	GE Healthcare, Buckinghamshire (UK)
Hybond N+ membrane	GE Healthcare, Buckinghamshire (UK)
Imperial Protein Stain	Pierce/Thermo Scientific, Rockford (USA)
Mini-PROTEAN 4-12% TGX gels	Bio-Rad, München (DE)
NucleoBond PC 10000	Machery&Nagel, Düren (DE)

NucleoBond Xtra Midi Plus	Machery&Nagel, Düren (DE)
NucleoSpin Plasmid	Machery&Nagel, Düren (DE)
NucleoSpin Extract II	Machery&Nagel, Düren (DE)
NuPAGE Antioxidant	Invitrogen, Karlsruhe (DE)
NuPAGE LDS Sample Buffer (4x)	Invitrogen, Karlsruhe (DE)
NuPAGE MOPS SDS Running Buffer (20x)	Invitrogen, Karlsruhe (DE)
NuPAGE Novex 4-12% Bis-Tris gels, 1mm	Invitrogen, Karlsruhe (DE)
NuPAGE Sample Reducing Agent (10x)	Invitrogen, Karlsruhe (DE)
SeeBlue Plus2 Protein Standard	Invitrogen, Karlsruhe (DE)

2.1.7 Commonly used buffers and solutions

10x DNA loading dye	30% [v/v] Glycerol 10 mM EDTA 0.25% [w/v] Bromophenol blue
10x PBS	1.37 M NaCl 27 mM KCl 100 mM Na ₂ HPO ₄ 20 mM KH ₂ PO ₄
1x PBS-T	1x PBS 0.1% [v/v] Tween-20
10x SDS running buffer	250 mM Tris base 1.92 M Glycine 1% [w/v] SDS
1 x SDS sample buffer	60 mM Tris-HCl (pH 6.8) 10% [v/v] Glycerol 2% [w/v] SDS 0.1% [w/v] Bromphenol blue 150 mM 2-Mercaptoethanol
10x TB	890 mM Tris base 890 mM Boric acid

10x TBE	890 mM Tris base 890 mM Boric acid 20 mM EDTA NaOH (pH 8.0)
1 x Transfer buffer	25 mM Tris base 0.192 M Glycine 0.1% [w/v] SDS 20% [v/v] Methanol
Chromatin loading buffer	10% [v/v] Glycerol 1x RB low
Colloidal Coomassie stain	0.08% [w/v] Coomassie Brilliant Blue G 250 20% [v/v] Methanol 1.6% [v/v] Ortho-Phosphoric acid 8% [w/v] Ammonium sulfate
Injection buffer	2% [w/v] SDS 100 mM NaCl 10 mM HEPES (pH 7.9) 50 mM Tris-HCl (pH 7.5) 2 mM MgCl ₂
MC buffer	10 mM HEPES (pH 7.6) 10 mM KOAc 0.5 mM Mg(OAc) ₂ 5 mM DTT
PD150 (pull-down wash buffer)	20 mM HEPES NaOH (pH 7.9) 10% [v/v] Glycerol 150 mM NaCl 0.1% [v/v] Triton X-100

RB High	10 mM Tris-HCl (pH 7.5) 1 mM EDTA NaOH (pH 8.0) 2 M NaCl 1 mM DTT
RB Low	10 mM Tris-HCl (pH 7.5) 1 mM EDTA NaOH (pH 8.0) 12.5 mM NaCl 1 mM DTT
Röder C	25% [v/v] Glycerol 20 mM HEPES (pH 7.9) 420 mM NaCl 1.5 mM MgCl ₂ 0.2 mM EDTA NaOH (pH 8.0)
Röder D	10% [v/v] Glycerol 20 mM HEPES (pH 7.9) 100 mM KCl 1.5 mM MgCl ₂ 0.2 mM EDTA NaOH (pH 8.0) 0.5 mM DTT 0.5 mM PMSF
SAU200	7 M Urea, 20 mM Sodium acetate (pH 5.2) 1 mM EDTA 200 mM NaCl 2 mM DTT
SAU600	7 M Urea 20 mM Sodium acetate (pH 5.2) 1 mM EDTA 600 mM NaCl 2 mM DTT

TW buffer	50 mM Tris-HCl (pH 7.5) 100 mM NaCl 1 mM EDTA 1% [v/v] Triton X-100 2 mM DTT
Unfolding buffer	7 M Guanidinium-HCl 20 mM Tris-HCl (pH 7.5) 10 mM DTT
Wash buffer	50 mM Tris-HCl (pH 7.5) 100 mM NaCl 1 mM EDTA 1 mM PMSF 1 mM Benzamidine 2 mM DTT
XL-SEC buffer	0.2% [w/v] SDS 100 mM NaCl 10 mM HEPES (pH 7.9)
1 M Tris-HCl buffer	1 M Tris base, desired pH adjusted with 37% [w/w] HCl
1 M HEPES buffer	1 M HEPES, desired pH adjusted with 5 M NaOH or 5 M KOH

2.1.8 Cell lines

HeLa S3 (human cervical cancer, Computer cell culture centre, BE) were provided by the facility for Cell Production, Max Planck Institute for Biophysical Chemistry, Göttingen, DE.

2.1.9 Chemically competent *Escherichia coli* strains

<i>DH5α</i> TM	Invitrogen, Karlsruhe (DE)
<i>dam-/dcm-</i>	New England Biolabs, Ipswich (USA)
BL21-CodonPlus (DE3)-RIL	Stratagene, La Jolla (USA)

2.1.10 Plasmids

Table 2.1 Plasmids used for chromatin reconstitution and protein expression.

Name	Promotor	Resistance	Supplier
pET3a_H2A/H2B/H3/H4	T7	Ampicillin	Karolin Luger, Colorado State Univ., Fort Collins (USA) (Dyer et al., 2004)
pET3a_H3A21C Δ1-20	T7	Ampicillin	Wolfgang Fischle, MPI for Biophysical Chemistry Göttingen (DE)
pET3a_H4R23C Δ1-22	T7	Ampicillin	Wolfgang Fischle, MPI for Biophysical Chemistry Göttingen (DE)
pET3d_H3K27C C110A	T7	Ampicillin	Wolfgang Fischle, MPI for Biophysical Chemistry Göttingen (DE)
pUC18_12x200-601	T7	Ampicillin	Daniela Rhodes, MRC Cambridge (UK) (Huynh et al., 2005)

2.1.11 Enzymes, Proteins and inhibitors

Antarctic phosphatase	New England Biolabs, Ipswich (USA)
Benzonase	Calbiochem, Darmstadt (DE)
M.SssI CpG methyltransferase	New England Biolabs, Ipswich (USA)
Proteinase Inhibitor Cocktail Complete, EDTA free	Roche, Mannheim (DE)
Restriction endonuclease enzymes	New England Biolabs, Ipswich (USA)
T4 DNA ligase	New England Biolabs, Ipswich (USA)
Taq polymerase	Chromatin Biochemistry (MPIbpc) (DE)
Trypsin	Roche, Mannheim (DE)
Trypsin	Promega, Mannheim (DE)

2.1.12 Antibodies

Table 2.2 Antibodies and antisera used for western blot analysis.

Name	Host	Supplier	Dilution
α -H3	rabbit, polyclonal	Abcam, ab1791	1 : 10000
α -Sept9	rabbit, polyclonal	William S. Trimble, SickKids Toronto (CA) (Surka et al., 2002)	1 : 250
α -Sept2	rabbit, polyclonal	William S. Trimble, SickKids Toronto (CA) (Xie et al., 1999)	1 : 250
α -SF3A3	rabbit, polyclonal	Reinhard Lührmann, MPI for Biophysical Chemistry Göttingen (DE) (Sharma et al., 2014)	1 : 2000
α -Bdp1	rabbit, polyclonal	Robert White, University of York (UK) (Fairley et al., 2003)	1 : 500
α -rabbit HRP	swine, polyclonal	Dako, P0399	1 : 5000

2.1.13 Antibiotics

Antibiotics:

Final concentrations used:

Ampicillin	100 μ g/ml
Kanamycin	50 μ g/ml
Chloramphenicol	34 μ g/ml
Tetracycline	12.5 μ g/ml

2.1.14 Oligonucleotides

Oligonucleotides were purchased from SIGMA (Steinheim, DE) and or were kindly provided by Alexandra Stützer (Chromatin Biochemistry Group, Max Planck Institute for Biophysical Chemistry, Göttingen, DE).

- Primers for PCR amplification of scavenger DNA for chromatin array reconstitution:
Forward 5' GTTATCCGCTCACAATTCCACACAACATAC 3'
Reverse 5' TAATGCAGCTGGCACGACAGGTTTC 3'

- Oligonucleotides for biotin labeling of 12x200-601 template for chromatin array reconstitution:

EcoRI_3'P 5'-GGGGGGGGATCCGGGGGGGp-3'

EcoRI_5'P_3'bio 5'-pAATCCCCCCCCGATCCCCCCC-biotin-3'

2.1.15 Peptides

Table 2.3 Peptides used for native chemical ligation.

Protein	Sequence	C-terminal	Supplier
H3K9me3	ARTKQTARK _{me3} STGGKAPRKQL	α-thioester	Petra Henklein, Charite Berlin (DE)
H3K9me2	ARTKQTARK _{me2} STGGKAPRKQL	α-thioester	Petra Henklein, Charite Berlin (DE)
H3K9me1	ARTKQTARK _{me1} STGGKAPRKQL	α-thioester	Petra Henklein, Charite Berlin (DE)
H4K20me3	SGRGKGGKGLGKGGAKRHRK _{me3} VL	α-thioester	Petra Henklein, Charite Berlin (DE)
H4K20me1	SGRGKGGKGLGKGGAKRHRK _{me1} VL	α-thioester	Petra Henklein, Charite Berlin (DE)
H4R3me2	SGR _{me2sym} GKGGKGLGKGGAKRHRKVL	α-thioester	Petra Henklein, Charite Berlin (DE)

2.1.16 Software

Adobe Creative Suite 4	Adobe Systems, San Jose (USA)
Cytoscape 3.3.0	Cytoscape Consortium (Shannon et al., 2003)
DAVID Bioinformatics Resources 6.6	National Institute of Allergy and Infectious Diseases (Huang da et al., 2009a, b)
MaxQuant and Andromeda	Max Planck Institute for Biochemistry (Cox and Mann, 2008; Cox et al., 2011)
Microsoft Office	Microsoft Corporation, Redmont (USA)
pLink	Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge (USA) (Purcell et al., 2007)
ProMass Deconvolution	Thermo Scientific (USA)

Proteome Discover	Thermo Scientific (USA)
R language for statistical computing	R Foundation for Statistical Computing (R, 2011)
Xlink Analyzer	(Kosinski et al., 2015; Pettersen et al., 2004)

2.2 Molecular biology methods

2.2.1 Transformation of chemically competent bacteria

A minimum of 10 ng of the respective plasmids were added to chemical competent cells and incubated for 30 min on ice. The transformation took place by a heat shock that was introduced by incubating the cell/plasmid mixture for 30 s at 42 °C. Following, the sample was incubated for 2 min on ice and 500 µl LB-medium were added. The cell suspension was shaken at 37 °C for 60 min before 100 µl were plated on antibiotic (2.1.13) containing LB agar plates. The agar plates were incubated overnight at 37 °C.

2.2.2 Analysis of nucleic acids

2.2.2.1 Concentration determination of nucleic acids

The concentration of DNA was determined by measuring the absorbance against a reference at 260 nm using the nanodrop spectrophotometer NanoDrop ND-1000. The DNA concentrations were calculated with the following equation: $1 \text{ OD}_{260\text{nm}} = 50 \text{ µg/ml double-stranded DNA}$ (Sambrook, 2001).

2.2.2.2 Nucleic acid gel electrophoresis

DNA fragments were separated according to their size on 0.5% to 2% [w/v] agarose gels in a TBE (or TB) buffer system (Sambrook, 2001). In general samples containing 300 ng of DNA and 1 x loading dye were loaded onto an agarose gel. Using a horizontal DNA electrophoresis cell (Sub-Cell-GT) 90 V to 120 V were applied for approximately 60 min. As a size reference 5 µl of a 1 kb Plus DNA Ladder run with all samples in parallel. Subsequently, the DNA was stained with 0.5 µg/ml ethidium bromide for 20 min and visualized on the ChemiDoc MP imaging system.

2.2.3 DNA precipitation

2.2.3.1 Ethanol

DNA was incubated with 0.3 M sodium acetate (pH 5.2) and 3 volumes ice cold ethanol (99% [v/v]) overnight at -20 °C (Sambrook, 2001). Afterwards, the mixture was centrifuged at 16000 x g for 30 min at 4 °C. The pellet was washed with 70% [v/v] ethanol, air dried and dissolved in double-distilled water (ddH₂O).

2.2.3.2 Isopropanol

DNA was mixed with 0.3 M sodium acetate (pH 5.2) and 0.7 volumes isopropanol and centrifuged immediately at 16000 x g for 30 min at 4 °C. The pellet was washed with 70% [v/v] ethanol, air dried and dissolved in ddH₂O (Sambrook, 2001).

2.2.3.3 PEG-6000

In order to separate large DNA fragments from smaller ones, DNA was sequentially precipitated using PEG (Lis and Schleif, 1975). For this, DNA was incubated with 0.5 M NaCl and 4 % [v/v] PEG-6000 for 10 min at RT. The mixture was centrifuged at 16000 x g for 15 min at 4 °C. The supernatant was collected and the PEG-6000 concentration was increased in 1% steps up to 9% [v/v] by repeating the centrifugation step of the resulting supernatants. The pellets resulting from each fraction were washed with ice cold 70% [v/v] ethanol, air dried at RT and dissolved in ddH₂O.

2.2.4 Preparation of DNA templates for chromatin array assembly

2.2.4.1 Preparation of biotinylated 12x200-601 DNA template

For amplification of pUC18_12x200-601 (2.1.10) the plasmid was transformed into chemically competent *dam-/dcm-* *E. coli* cells (2.1.9). 200 ml ampicillin containing LB medium were inoculated with a single bacteria colony at 37 °C and 120 rpm overnight. For large scale preparation, 2.5 l of LB medium supplemented with ampicillin were inoculated with 5 ml of the overnight culture. This culture was grown at 37 °C overnight, shaking at 120 rpm. Cells were

harvested by centrifugation for 15 min at 6000 x g at 4 °C. Cell lysis and DNA purification were done using the NucleoBond PC 10000 kit according to the manufacturer's instructions.

The 12x200-601 sequence was obtained by digestion of purified plasmid DNA using the restriction endonuclease enzymes *BsrI*, *EcoRI*, *HaeII*, and *DdeI*. For the digest, 300 U of each enzyme per mg of DNA were used. The reaction took place in 1 x NEBuffer 2.1 (New England Biolabs) at 37 °C overnight. After inactivation of the restriction endonuclease enzymes (at 80°C for 20 min) the 2.4 kb 12x200-601 DNA template was separated by PEG precipitation (2.2.3.3).

The 12x200-601 DNA was biotinylated by ligation of a biotinylated ds oligonucleotide to the 5' *EcoRI* overhang of the DNA. An equimolar mix of the "*EcoRI*_3'P" and "*EcoRI*_5'P_3'bio" oligonucleotides were annealed by heating them to 95 °C and allowing them to cool down to RT. Prior ligation, the 12x200-601 DNA was dephosphorylated using 1 U Antarctic phosphatase per 2 µg DNA according to the manufacturer's instructions. The ligation was performed using a tenfold molar excess of the (ds) oligonucleotide. The reaction was carried out with 2.5 U T4 DNA ligase per 1 µg of DNA for at least three hours at RT. The biotinylated DNA was extracted by addition of an equal volume of phenol:chlorophorm:isoamyl alcohol [25:24:1] in phase lock heavy tubes followed by isopropanol precipitation (2.2.3.2) of the aqueous phase.

2.2.4.2 Methylation of biotinylated 12x200-601DNA template

CpG sites of the 12x200-601 DNA template were methylated using the *M.SssI* methyltransferase. 500 µg of DNA were incubated with 240 U of *M.SssI* in 50 mM NaCl, 10 mM EDTA, 10 mM Tris-HCl, pH 7.9 and 160 µM SAM at 37 °C. After three hours, 80 µM SAM and 100 U of *M.SssI* were added, the reaction was left to proceed overnight at 37 °C. To ensure complete methylation another 130 µM of SAM and 80 U of *M.SssI* were added and the reaction was incubated for four more hours. The DNA was extracted with PCI [25:24:1] and precipitated using ethanol (2.2.3.1).

The efficiency of methylation was tested using restriction endonucleases unable to cleave at methylated CpG sites, like *BsaAI*, according to the manufacturer's instructions.

2.2.4.3 Preparation of scavenger DNA

The scavenger DNA is a 147 bp sequence of the backbone of the pUC18_12x200-601 plasmid (2.1.10) that does not contain the 200-601 nucleosome positioning sequence. The DNA was amplified by polymerase chain reaction (PCR) according to the following PCR mix and PCR cycle program:

<u>PCR mix (100 μl):</u>		<u>PCR cycle program:</u>	
		Temperature	Time
100 mM	Tris-HCl pH 8.8	95 °C	2 min
500 mM	KCl	95 °C	30 s
0.8% [v/v]	NP-40	95 °C	30 s
0.2 mM (each)	dNTP mix	68 °C	30 s
0.2 μ M (each)	primers	72 °C	30 s
1.5 mM	MgCl ₂	72 °C	1 min
50 ng	DNA template	10 °C	∞
20 U/ml	Taq polymerase		

The PCR product was extracted with PCI [25:24:1] and subsequently precipitated with isopropanol (2.2.3.2).

2.3 Protein biochemistry methods

2.3.1 Detection and analysis of proteins

2.3.1.1 Concentration determination of proteins

The concentration of histone proteins was determined by UV adsorption at 276 nm on a NanoDrop ND-1000 using published molar extinction coefficients listed in table 2.4 (Luger et al., 1999). The protein concentration of nuclear extracts was measured by UV adsorption at 280 nm assuming 1 OD₂₈₀ = 1 mg/ml.

Table 2.4 Values used for the determination of histone concentrations.

Protein	Molecular weight (Da)	Molar extinction coefficient (M⁻¹cm⁻¹)
H2A	13960	4050
H2B	13774	6070
H3	15273	4040
H4	11236	5400

2.3.1.2 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated and visualized by discontinuous polyacrylamide gel electrophoresis in presence of SDS based on the method of Laemmli (Laemmli, 1970) according to standard protocols (Gallagher, 2006). 15% or 18% Tris-glycine resolving gels in combination with a 4% stacking gel were used for the analysis of histone proteins. The gels were cast and run using the Mini-PROTEAN electrophoresis system according to instructions provided by the manufacturer. Before loading the protein samples onto the gel, they were boiled for 5 min at 95 °C in SDS sample buffer. As a reference 5 µl of SeeBlue Plus2 pre-stained protein standard was used. Protein samples were run in 1 x SDS running buffer at constant current at 15 mA (250 V maximum) per gel until the bromophenol blue front reached the bottom of the gel.

The separation of complex protein mixtures, including samples for mass spectrometry analysis, was performed using NuPAGE Novex 4-12% Bis-Tris gradient gels in combination with the NuPAGE Pre-cast system according to the manufacturer instructions. For all runs 1 x MOPS buffer, antioxidant, LDS sample buffer and reducing agent were used.

2.3.1.3 Coomassie staining of protein SDS-PAGE gels

Proteins separated on SDS-PAGE gels were either stained overnight with Colloidal Coomassie stain solution or for 10 to 20 min with Imperial Protein Stain. Gels were destained with ddH₂O for at least four hours.

2.3.1.4 Silver staining of protein SDS-PAGE gels

Silver staining was performed according to the protocol of Blum et al. (Blum H., 1987). For silver staining of proteins separated on SDS-PAGE gels all incubation steps took place at RT on a shaker. Gels were incubated in fixative solution (50% MeOH, 12% HOAc) overnight. Afterwards, they were incubated twice in 50% EtOH followed by an incubation in 30% EtOH for 20 min each. Gels were sensitized in 0.8 mM sodium thiosulfate for exactly 60 s. Sodium thiosulfate solution was removed by three wash steps with ddH₂O, 20 sec each time. Subsequently, the gels were incubated in silver nitrate solution (2 g/l silver nitrate, 0.026% formaldehyde) for 20 min. Before development the gels were washed three times with ddH₂O. The water was replaced by developer solution (60 g/l Na₂CO₃, 0.0185% formaldehyde, 16 µM sodium thiosulfate) until protein bands on the gel turned brown. The process was terminated by addition of fixative solution.

2.3.1.5 Western blot

Immunoblotting of protein samples separated by SDS-PAGE was done using the MiniTrans Blot system. Proteins were transferred onto nitrocellulose membranes. For this, the membrane was directly laid onto the SDS-PAGE gel and both were sandwiched between three layers of Whatman paper equilibrated in 1 x transfer buffer. Proteins were blotted onto the membrane in 1 x transfer buffer for one hour at 100 V at 4 °C. The membrane was rinsed with 1 x PBS-T and subsequently blocked with 5% milk for 30 min at RT. Afterwards, the membrane was incubated with a primary antibody overnight at 4 °C. All antibodies were diluted in 5% milk according to table 2.2. In order to remove non-bound antibodies, the membrane was washed three times with 1 x PBS-T for 10 minutes each. The membrane was then incubated with a secondary antibody at RT for at least two hours, followed by three wash steps as described above. The detection was done using an enhanced chemiluminescence system (ECL) according to the manufacturer's instructions. The imaging was done using the ChemiDoc MP System.

2.3.2 Recombinant proteins

2.3.2.1 Expression and purification of recombinant histone proteins

Purification of *Xenopus leavis* histones was performed as described before (Luger et al., 1999) with certain alterations.

The respective plasmids (2.1.10) were transformed into BL21-CodonPlus (DE3)-RIL *E. coli* cells. 200 ml of ampicillin containing 2xYT medium were inoculated with a single bacteria colony overnight at 37 °C and shaking at 120 rpm. The overnight culture was used to inoculate 2xYT medium (1:1000) that was incubated at 37 °C and shaking at 120 rpm. When the culture reached an $OD_{600} = 0.5$ expression was induced with 0.2 mM IPTG and the culture was incubated for four more hour. Afterwards, cells were harvested by centrifugation at 6000 x g for 15 min at 4 °C.

In order to purify the histone proteins inclusion bodies were isolated. All steps were performed at 4 °C. The cell pellet resulting from 2 l bacteria culture was resuspended in 100 ml Wash buffer and cells were lysed using a Canadian Press. After 4 cycles the suspension was centrifuged for 20 min at 23000 x g. The pellet was resuspended in 100 ml TW buffer and centrifuged for 20 min at 23000 x g. The wash step was repeated twice. In order to remove Triton X-100 the inclusion bodies were additionally washed two times with Wash buffer.

To extract histone proteins from inclusion bodies the pellet was soaked with 1 ml DMSO, minced and incubated at RT for 30 min. Next, 30 ml of Unfolding buffer were added. The suspension

was stirred for 1 h at RT and then centrifuged at 23000 x g for 20 min. The resulting pellet was re-extracted by adding another 10 ml of Unfolding buffer for 30 min. After another centrifugation step both supernatants were pooled and dialyzed three times against 2 l of SAU200.

For purification of the histone proteins a Q-Sepharose XK26/20 and a SP-Sepharose XK26/20 column were connected in series and equilibrated with SAU200 before the extracted proteins were loaded onto the column combination. After the input has passed the Q-Sepharose column was disconnected and washing was continued only with the SP-Sepharose column. Histone proteins were eluted applying a linear gradient from 0% to 100% of SAU600 over one column volume. Histone containing elution fractions were pooled, dialyzed three times against 2 mM DTT, lyophilized and stored at – 80 °C.

For H4R23C Δ 1-22 additional alterations to the protocol were implemented. The expression of the protein was not induced before the bacteria culture reached an $OD_{600} = 1$ and induction took place for one hour. Protein extraction from inclusion bodies was extended to two hours. H4R23C Δ 1-22 proteins were eluted from the SP column applying a step gradient that was first set to 20% of SAU600, until the UV absorption at 276 nm reached again the value that was detected when the column was equilibrated, followed by a gradient from 20% to 100% of SAU600 over one column volume. The elution of proteins was visible by two peaks monitored at 276 nm. Elution fractions of the first peak were kept whereas the elution fractions of the second peak were pooled, diluted 1:3 with SAU0 (the same as SAU200 but without NaCl) and reloaded onto the newly equilibrated SP column in a second run. H4R23C Δ 1-22 was again eluted with 20% SAU600. Histone containing elution fractions of the first and the second run were pooled, dialyzed three times against 2 mM DTT, lyophilized and stored at – 80 °C.

2.3.2.2 Introduction of posttranslational modifications by native chemical ligation

The introduction of methylation sites at the N-termini of histone H3 and histone H4 was achieved by native chemical ligation as described before (Shogren-Knaak and Peterson, 2004) with a few alterations. Synthetic N-terminal histone polypeptides containing modified amino acid residues were ligated to the truncated histones H3A21C Δ 1-20 or H4R23C Δ 1-22, respectively. The reaction took place using a molar histone:peptide ratio of 1:5 with a final peptide concentration of 2 mM.

Briefly, the respective amount of histone proteins were dissolved in ligation buffer (6 M guanidinium-HCl, 100 mM KP_i pH 7.9, 20 mM TECEP, 50 mM MPAA) and added to the lyophilized peptides. The reaction mixture was incubated at RT for at least 40 h and finally dialyzed three times against SAU200. The reaction mixture was loaded onto a HiTrap SP HP 1ml column equilibrated with SAU200. The histone proteins were eluted by applying a linear

gradient from 0% to 100% of SAU600 over 20 column volumes. Fractions containing full length histone proteins were pooled, dialyzed three times against 2 mM DTT, lyophilized and stored at – 80 °C. The progress of the reaction was monitored by SDS-PAGE and confirmed by electrospray mass spectrometry.

2.3.2.3 Introduction of methyl lysine analogs

The introduction of methyl lysine analogs (MLA) was achieved by performing an alkylation reaction as described before (Simon, 2010). All methylation states of H3K27 were produced using the histone mutant H3K27C, C110A that was dissolved in Alkylation buffer (4 M guanidine hydrochloride, 1 M HEPES pH 7.8, 10 mM DL-methionine). All alkylation reactions were performed with 10 mg of H3K27C, C110A in a final volume of 1ml.

The introduction of the tri-methyl lysine analog was performed at 50 °C in presence of 400 mM (2-bromoethyl)-tri-methylammonium bromide and 20 mM DTT in the dark with occasional flicking. After 5 h the reaction was quenched with 50 µl 2-mercaptoethanol for 30 min at RT. The MLA histones were purified using a PD-10 column according to the manufacturer's instructions using 2 mM 2-mercaptoethanol for equilibration and elution.

The introduction of the di-methyl lysine analog was performed at RT in the presence of 50 mM (2-chloroethyl)-di-methylammonium chloride and 20 mM DTT in the dark with occasional flicking. After 2.5 h additional 50 mM of (2-chloroethyl)-di-methylammonium chloride were added to the reaction mixture. After a total of 4.5 h the reaction was quenched with 50 µl 2-mercaptoethanol for 30 min at RT. The MLA histones were purified using a PD-10 column as described for the tri-methyl lysine analog.

The introduction of the mono-methyl lysine analog was performed at RT in the presence of 300 mM (2-chloroethyl)-methylammonium chloride and 20 mM DTT in the dark with occasional flicking. After 5 h the reaction was quenched with 50 µl 2-mercaptoethanol for 30 min at RT. The MLA histones were purified as described for the tri-methyl lysine analog.

2.3.3 Preparation of recombinant chromatin

2.3.3.1 Assembly of histone octamers

The assembly of octamers composed of the histone proteins H2A, H2B, H3 and H4 were performed according to a generalized protocol (Luger et al., 1999). Each of the lyophilized histone proteins were completely dissolved in Unfolding buffer to a concentration of 2 mg/ml. All four histone variants were mixed in equimolar ratios and adjusted to a final concentration of 1

mg/ml. The protein mixture was dialyzed three times against 2 l RB high at 4 °C. Afterwards, the assembled octamers were concentrated to a final volume of 50 µl using Amicon Ultra centrifugal filter devices and loaded onto a HiLoad 16/60 Superdex 200 prep grade gel filtration column installed on an ÄKTA FPLC system. The octamer-containing fractions were pooled and 50% [v/v] glycerol was added for long term storage at -20 °C.

2.3.3.2 Chromatin reconstitution by salt dialysis

The reconstitution of oligonucleosomal chromatin arrays composed of 12 nucleosomes was performed according to a previously established protocol by dialysis from high salt to low salt (Luger et al., 1999). The concentration of octamers in RB high buffer was determined by UV spectroscopy ($0.45 \text{ OD}_{276 \text{ nm}} = 1 \text{ mg/ml octamer} = 92.2 \text{ µM}$). The DNA template described at 2.2.4.1 and histone octamers (2.3.3.1) were mixed in a molar ratio of 1.1(-1.3):1 (octamer:DNA) in RB high buffer. Additionally, an equimolar amount of scavenger DNA (2.2.4.3) was added to bind excess octamer. The salt gradient from 2 M NaCl to 12.5 mM NaCl was achieved by continuously replacing RB high with RB low over 30 h using a peristaltic pump.

The reconstitution was analyzed by agarose gel electrophoresis on a 0.5% [w/v] gel in 0.2 x TB buffer at 4 °C. Furthermore, the saturation level of reconstituted chromatin arrays was analyzed by restriction digest using the *Ava*I restriction enzyme and analytical ultracentrifugation (2.3.4.2).

2.3.4 Molecular characterization of recombinant chromatin

2.3.4.1 *Ava*I digest

The saturation level of reconstituted chromatin arrays was analyzed by a digest with the restriction enzyme *Ava*I, which cuts within the 200-601 nucleosome positioning sequence of the chromatin DNA template (2.2.4.1) only when not occupied by histone octamers.

1 µg of chromatin arrays (2.3.3.2) and 0.5 µg of the DNA template (2.2.4.1) were incubated with 20 U of *Ava*I in a total volume of 30 µl for 3 h at 37 °C according to the manufacturer instructions. The digested samples were analyzed by agarose gel electrophoresis on a 2% [w/v] gel.

2.3.4.2 Analytical ultracentrifugation

For the analysis of oligonucleosomes by the sedimentation velocity method using a Beckmann XL-A centrifuge, 400 μ l containing 0.5 OD₂₆₀ oligonucleosomes were prepared in RB low buffer. The buffer sector of the cells was filled with 400 μ l RB low buffer, while the sample sector was filled with 390 μ l of the chromatin containing solution. A speed of 15.000 rpm was used for sedimenting the oligonucleosomes within a series of 25-35 scans. Data analysis was performed using the software UltraScan II (Demeler). In general, after assignment of meniscus, cell bottom and plateau for each of the analyzed cells, spikes and time-invariant noise as well as unsatisfactory scans (where a sigmoidal sedimentation curve cannot be assigned) were manually removed. Initial sedimentation distribution was calculated for the viscosity of RB low and the sample frictional coefficient. The initial result was used to perform van Holde-Weischet boundary analysis on each of the sedimentation scans. The resulting boundary analysis was used for plotting the distribution of corrected sedimentation coefficients.

2.4 Cell culture and metabolic labeling

2.4.1 Stable isotope labeling by amino acids in cell culture (SILAC) of HeLa S3 cells

HeLa S3 cells were cultivated as described before (Nikolov et al., 2011). The cells were grown in custom High Glucose Dulbecco's Modified Eagle's Medium (DMEM) lacking arginine and lysine. The medium was supplemented with either light arginine (Arg0) and lysine (Lys0) or heavy amino acids (Arg6 and Lys4) (2.1.5) to a final concentration equivalent to 50 mg/ml light amino acids. In addition, 1/10 volume of dialyzed fetal bovine serum and 1 x Penicillin/Streptomycin were added to the medium.

The culture was started by bringing 10^8 cells into a spinner flask containing 100 ml medium. Cells were grown at 37 °C, 5% CO₂ and 95% relative humidity. Over at least six passages the culture was expanded to 1.5 l and finally transferred to a 5 l bioreactor where cells were grown under standard conditions (2×10^6 cells/ml, continuously addition of synthetic air and dissolved oxygen level kept at pO₂ = 20 using a feedback monitoring system).

2.4.2 Preparation of nuclear extracts

The preparation of nuclear extracts was done as described before (Dignam et al., 1983). Generally, 10^{10} SILAC labeled HeLa S3 cells were harvested by centrifugation for 5 min at 2000 rpm. The pellet was washed twice with 1 x PBS and resuspended in 1.25 volumes MC buffer

supplemented with 1/500 volumes 0.25 M DTE and EDTA-free protease inhibitor cocktail according to the manufacturer's instructions. The suspension was incubated on ice for 5 min and homogenized using a 50 ml Dounce homogenizer, performing 18 strokes. Nuclei were pelleted by centrifugation at 18 000 x g for 5 min at 4 °C. The pellet was resuspended in 1.3 volumes Röder C buffer supplemented with 1/500 volumes 0.25 M DTE and 1/200 volumes 0.1 M PMSF. Nuclei were disrupted using a Dounce homogenizer, performing 20 strokes. The suspension was stirred for 40 min at 4 °C and finally centrifuged at 30 000 x g for 30 min at 4 °C. The supernatant was frozen in liquid nitrogen and stored at -80 °C.

Three nuclear extracts were prepared from individual cell cultures for each labeling, heavy and light labeled. The three extracts were pooled according to their labeling and dialyzed three times against 2 l of Röder D buffer and centrifuged at 9000 x g for 2 min at 4 °C. The supernatant was aliquoted, frozen in liquid nitrogen and stored at -80 °C.

2.5 Biochemical binding assays

2.5.1 Chromatin affinity purification

Chromatin affinity purification using differentially SILAC-labeled HeLa S3 nuclear extracts was essentially done as described before (Nikolov et al., 2011). All steps were performed at 4 °C. 40 µg of chromatin arrays of two different modification states were separately immobilized on 160 µl paramagnetic streptavidin coated beads in PD150 buffer supplemented with 1 mg/ml BSA overnight. Chromatin-bound beads were washed three times with 800 µl PD150 wash buffer. In parallel, 1 ml of heavy and light labeled nuclear extract was incubated with 40 µl paramagnetic streptavidin coated beads for 1h to remove unspecific binding proteins. 4 mg of nuclear extract were separated from beads and incubated with chromatin-bound beads. After four hours of incubation, beads were washed three times with PD150 wash buffer and proteins were eluted with 1 x NuPAGE LDS Sample buffer supplemented with 2 mM MgCl₂ and 10 mM Tris-HCL pH 7.9. Eluates from pull-downs done with heavy and light nuclear extracts were mixed in a 1:1 ratio. To ensure complete release of all chromatin-bound proteins and to improve the following separation step, the mixed sample was digested with 1250 U benzonase for 45 min at 37 °C. After addition of 1 x NuPAGE reducing agent the samples were heated at 70 °C for 10 min before loading onto a 4% - 12% SDS-PAGE gel. After separation, the gel was stained with Coomassie overnight.

2.5.2 Protein-protein cross-linking on beads

150 µg of chromatin arrays (2.3.3.2) were immobilized on 600 µl paramagnetic streptavidin coated beads in PD150 buffer overnight. Chromatin-bound beads were washed 3 times with 3 ml PD150 wash buffer and incubated with 2 ml pre-cleared nuclear extract (2.5.1) at 4 °C. After 3 h 20 µM BS3 was added to the sample, which was incubated for 4 h in total. Chromatin-bound beads were washed once with 3 ml PD150 wash buffer to remove non-specifically bound proteins. To chemically cross-link chromatin-bound proteins beads were incubated twice in 2 ml PD150 wash buffer supplemented with 200 µM BS3 for 10 min at RT. Cross-linked proteins and chromatin were eluted from the beads by adding 55 µl Injection buffer and 1250 U benzonase followed by incubation for 45 min at 37 °C. Afterwards the sample was heated to 70 °C, mixed several times using a vortex mixer, and finally separated from magnetic beads using a magnetic rack.

50 µl of the sample were injected onto a Superdex 200 column equilibrated with XL-SEC buffer. For separation, a flow rate of 55 µl/min was applied and 50 µl fractions were collected. Elution fractions were pooled according to the peaks of the elution profile monitored at 215 nm and analyzed by silver staining of SDS-PAGE gels and western blot analysis.

Proteins of the pooled fractions containing cross-linked material were precipitated using acetone at -20 °C overnight. Pellets were resuspended in 1% RapiGest before tryptic digestion (2.6.2). Subsequently, acetonitrile was added to the digested samples to a final proportion of 10% [v/v]. Following, the sample was injected onto a Superdex peptide column that was equilibrated with 30% acetonitrile and 0.1% formic acid. For separation, a flow rate of 55 µl/min was applied and 50 µl fractions were collected. Elution fractions were pooled according to the peaks of the elution profile monitored at 215 nm and subjected to mass spectrometric analysis (2.6.4).

2.6 Mass spectrometry methods

2.6.1 In-gel proteolysis of proteins and peptide extraction

In-gel digest of proteins and the following extraction of peptides were essentially done as described before (Shevchenko et al., 2006; Shevchenko et al., 1996). All steps of the in-gel digest were performed at 25 °C under shaking at 1050 rpm. Each SDS-PAGE gel lane was sliced into either 11 or 22 equidistant gel pieces regardless of staining. In order to reduce proteins embedded in gel, slices were dehydrated with acetonitrile and rehydrated with 100 mM dithiothreitol (DTT) in 50 mM ammonium hydrogen carbonate (NH_4HCO_3) pH 8.0 and incubated at 56 °C for 50 min. After dehydrating the gel slices in acetonitrile, reduced cysteine residues

were alkylated in 60 mM iodoacetamide, 50 mM NH_4HCO_3 pH 8.0 for 20 min at 26 °C in the dark. Gel pieces were washed afterwards with 50 mM NH_4HCO_3 pH 8.0, which was replaced by acetonitrile to completely dehydrate them. Finally, gel pieces were rehydrated by incubation in 20 μl digestion buffer (12.5 ng/ μl trypsin (Roche), 41.7 mM NH_4HCO_3 pH 8.0, 4.2 mM CaCl_2) for 40 min on ice. 60 μl digestion buffer without trypsin were added to the gel pieces and incubated at 37 °C overnight.

For the extraction of peptides from gel pieces incubation steps were performed at 37 °C, shaking at 1050 rpm for 15 min. 60 μl ddH₂O and 160 μl acetonitrile were added to the digested samples. Once gel pieces were shrank, the supernatants were collected and 100 μl 5% formic acid were added to the dehydrated gel pieces. Subsequently, gel pieces were dehydrated again by adding acetonitrile and the supernatants were collected and pooled with those from the first extraction step. Supernatants were evaporated to dryness in a centrifugal concentrator. Dried peptides were kept at -20 °C or were immediately dissolved in loading buffer (5% [v/v] acetonitrile, 1% [v/v] formic acid) and analyzed by LC-MS/MS (2.6.3).

2.6.2 In-solution proteolysis of proteins

For the in-solution digest of proteins all reagents used were dissolved in 25 mM ammonium bicarbonate pH 8.5. Proteins precipitated with acetone were dried in a centrifugal concentrator. The pellet was dissolved in 5 μl 1% RapiGest (sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate) (Yu et al., 2003). In order to reduce the proteins 5 μl 50 mM DTT were added and the sample was incubated for 1 h at 37 °C. Subsequently, 5 μl 100 mM iodoacetamide were added and the sample was incubated for 1 h at 37 °C in the dark. The concentration of RapiGest was decreased to 0.1% by adding 35 μl trypsin (Promega) solution corresponding to an enzyme:protein ratio of approximately 1:20 enzyme-to-substrate ratio (w:w). The digest took place at 37 °C overnight. To decompose the RapiGest the sample was supplemented with 5 μl 40% TFA and incubated at 37 °C for two more hours. Finally, the sample was centrifuged at 16 000 x g for 30 min and the supernatant was transferred to a new tube. Tryptic peptides were dried in a centrifugal concentrator, dissolved in loading buffer (5% [v/v] acetonitrile, 1% [v/v] formic acid) immediately afterwards and analyzed by LC-MS/MS (2.6.4).

2.6.3 LC-MS/MS analysis of peptides on LTQ-Orbitrap

Tryptic peptides from in-gel digestion (2.6.1) were analyzed by nanoflow chromatography coupled to hybrid ion trap-orbitrap mass spectrometry using a vented column setup (Licklider et al., 2002). Dried peptides were dissolved in 20 μl loading buffer (5% [v/v] acetonitrile, 1% [v/v]

formic acid) and 20% of the sample was injected and enriched on a custom-packed reversed-phase C18 precolumn, followed by separation on a custom-packed reversed-phase C18 column (0.075 mm ID x 150 mm, Reprosil-Pur 120 C18-AQ, 3 μ m) using a 90 min linear gradient of 5-35% [v/v] acetonitrile, 0.1% [v/v] formic acid at 300 nl/min. The eluent was analyzed in positive ion mode on a LTQ-Orbitrap Velos hybrid quadrupole/orbitrap mass spectrometer equipped with a FlexIon nanoSpray source with a stainless steel emitter and operated under Excalibur software 2.1.0.1140 to perform data-dependent acquisition. Each experimental cycle consisted of one full MS scan across the 350-1600 m/z range that was acquired at a resolution setting of 30,000 FWHM, an automatic gain control target of $10e^6$ and lock mass correction for m/z 441.120025 from polysiloxane. Up to the 15 most abundant peptide precursors of charge states 2 to 4 above an intensity threshold of $2 \times 10e^3$ were selected at an isolation width of 2.0 Th, fragmented by collision-induced dissociation (CID) using helium as collision gas at a normalized collision energy setting of 37.5% and an activation time of 10 ms. The product ion spectra were recorded using an AGC target of $2 \times 10e^5$ and a maximum fill time of 100 ms. Precursor m/z values selected for MS/MS were excluded for 30 s after one occurrence. Three technical replicates were acquired per sample.

In-gel digestions resulted in eleven samples per chromatin affinity purification experiment were measured on a LTQ-Orbitrap Velos using a 90 min total gradient, in-gel digestions resulted in 22 samples were analyzed on a LTQ-Orbitrap XL mass spectrometer using a 50 min gradient.

2.6.4 LC-MS/MS analysis of peptides on QExactive

Tryptic in-solution digested peptides (2.6.2) were analyzed by nanoflow chromatography coupled to a hybrid quadrupole-orbitrap mass spectrometer using a vented column setup (Licklider et al., 2002). Samples were injected and enriched on a custom-packed reversed phase-C18 precolumn, followed by separation on a custom-packed reversed-phase C18 column (0.075 mm ID x 150 mm, Reprosil-Pur 120 C18-AQ, 3 μ m) using a 70 min total gradient of 5-40% [v/v] acetonitrile, 0.1% [v/v] formic acid at 300 nl/min. The eluent was analyzed in positive ion mode on a QExactive hybrid quadrupole-orbitrap mass spectrometer equipped with a FlexIon nanoSpray source with a stainless steel emitter and operated under Excalibur software 2.2.0.48 to perform data-dependent acquisition. Each experimental cycle consisted of one full MS scan across the 350-1600 m/z range that was acquired at a resolution setting of 70,000 FWHM, an automatic gain control target of $10e^6$ and lock mass correction for m/z 441.120025 from polysiloxane. Up to the 15 most abundant peptide precursors of charge states 3 to 7 above an intensity threshold of $2 \times 10e^4$ were selected at an isolation width of 2.0 Th, fragmented by higher energy collisional dissociation (HCD) using nitrogen as collision gas at a normalized collision energy setting of 25%, and their product ion spectra recorded using an AGC target of $2 \times 10e^5$

and a maximum fill time of 60 ms. Precursor m/z values selected for MS/MS were excluded for 30 s after one occurrence.

2.6.5 Molecular weight determination of histone proteins

Histone proteins were dissolved in 20% [v/v] acetonitrile and 0.1% [v/v] formic acid and injected to a LTQ XL Linear Ion Trap mass spectrometer. The proteins were analyzed in a scan mode ranging from 500-2000 m/z . The spectra were extracted and deconvoluted using the Small Protein mode of the ProMass Deconvolution software.

2.7 Raw data processing and data analysis

2.7.1 Mass spectrometric raw data processing with MaxQuant

Protein identification and quantification of raw data files derived from mass spectrometric measurements (2.6.3, 2.6.4) were performed using the MaxQuant software, version 1.5.2.8 (Cox and Mann, 2008; Cox et al., 2011). Enzyme specificity was set to trypsin/P at a maximum of two missed cleavages, methionine oxidation was set as variable and cysteine carbamidomethylation was set as fixed modification. All other parameters were set according to the default settings with the following exceptions. For experiments done with SILAC-labeled nuclear extracts $^{13}\text{C}_6$ -arginine and $^{13}\text{C}_4$ -lysine were set as fixed modifications. Additionally, the minimum length of peptides was decreased to 6 and the maximal number of modification per peptide to 4. Proteins were identified using the Andromeda algorithm searching against the UniProtKB SwissProt Human protein sequence database 2014.01 (134921 protein entries) supplemented with the MaxQuant common contaminants database (248 entries).

2.7.2 Mass spectrometric raw data processing with pLink

In order to identify cross-linked peptides, raw files obtained from mass spectrometric measurements (2.6.4) were transformed into mgf files using Proteome Discover and analyzed by the pLink software version 1.07 (Purcell et al., 2007). Raw files of one experiment were analyzed in parallel and results were summarized in one output table. The *ini* file used for the data search was configured in the following way. Carbamidomethyl was set as a fixed modification, oxidation, pyro-glutamic acid and N-terminal acetylation were set as variable modifications, the search was done in mode 1, noninterexport was set as true and the *fdr* was set to 0.01. All other settings

were defined according to the experimental design and/or correspond to the default setting of the example *ini* file downloaded with the software. Raw files used for the pLink search were also analyzed by MaxQuant (2.7.1). Identified proteins summarized in the proteinGroups.txt of MaxQuant were used as the reference database for the pLink search. The reference database included only identified proteins with more than 2 “razor and unique peptides”. Additionally, contaminants and reverse hits were excluded. The amino acid sequences of identified histone proteins were exchanged by those from *Xenopus laevis* (2.1.10). The most abundant proteins of the samples were determined by dividing the number of identified “unique and razor peptides” by the molecular weight of the corresponding proteins. The most abundant proteins were arranged according to the quotient starting with the highest value.

In total, three pLink searches using different databases were performed for each experiment. The first database included the top 100 of the most abundant proteins, the second database was composed of the top 150 most abundant proteins and the third database included all identified proteins of each experiment, respectively. For spectra that were annotated in more than one of the pLink searches the spectrum with the highest score was chosen for further analysis. Those spectra that were annotated in more than one pLink search displaying different outcomes were excluded from further analysis.

Visualization of identified protein-protein cross-links was done using xiNET (Combe et al., 2015). Visualization and analysis of cross-linking data in the context of their three-dimensional protein structures were done using the Xlink Analyzer software (Kosinski et al., 2015; Pettersen et al., 2004).

2.7.3 Data filtering and visualization of quantified MS data

The MaxQuant output table proteinGroups.txt was imported to the statistical software package PERSEUS (Cox and Mann, 2008), version 1.5.0.26. Normalized heavy-to-light ratios (H/L) and intensities were extracted and log2-transformed. Only proteins with a ratio count of at least 3 in forward and reverse experiments for a given modification state were considered for quantification. After removal of reverse and contaminant entries, the statistical significance of individual quantification results was assessed using Perseus’ “significance A” algorithm using a two-sided False Discovery Rate estimation to a p value of 0.1 and application of a Benjamini-Hochberg correction (Cox and Mann, 2008). The processed data were visualized using the open-source statistical software package R (version 3.2.3). For data visualization only, missing values in one of the experiments were imputed to zero.

2.7.4 Protein-protein interaction analysis

Protein-protein interaction analyses were done using the STRING database (Jensen et al., 2009). Leading UniProt IDs reported in the proteinGroups.txt of MaxQuant (2.7.1) were used to identify known and predicted functional and structural interactions (default settings) for the design of protein-protein association networks.

2.7.5 Annotation enrichment analysis

Functional annotation enrichment analyses were done using the functional annotation chart of the online software package DAVID (Huang da et al., 2009a, b). Results were obtained by using default settings of the statistical parameters with the exception of the EASE score (enrichment) cutoff, which was set to 0.05. All genome genes of *Homo sapiens* were used as the background database. The background database was provided by DAVID (Huang da et al., 2009a, b). Terms were annotated to significant enriched and significant depleted proteins of a certain chromatin modification state, respectively. Only enriched annotated terms were included in further analyses. The processed data were visualized using the open-source statistical software package R (version 3.2.3).

3 Results

Nucleosomal arrays have been successfully used to characterize the diversity and complexity of PTM-regulated protein interactomes associating with chromatin (Nikolov et al., 2011). Nevertheless, comprehensive information regarding the impacted of heterochromatic chromatin modifications on proteins binding specifically to chromatin is still lacking. The current information content rarely allows drawing conclusions about the extent of functional outcomes associated with certain histone modification patterns.

In this study, chromatin affinity purification (ChAP) coupled to quantitative mass spectrometry was optimized predominately in terms of statistical evaluation, to investigate the interactomes brought about by ten different chromatin modifications. Furthermore, the method was extended to not only identify the interactomes in the context of singly-modified chromatin arrays but also the combination of two posttranslational histone modifications. Thus, the extended workflow enabled the investigation of the impact of histone PTMs on protein binding to chromatin and histone modification crosstalk on a global level. Additionally, a workflow combining ChAP, chemical cross-linking and mass spectrometry is introduced. The new combination of experimental steps enabled the identification of site-specific protein-protein interactions. Moreover the assignment of a network on the basis of these physical protein interactions provided information about the hierarchy of protein binding to chromatin.

3.1 12mer nucleosomal arrays are of reproducible quality independent of incorporated modifications

In order to expand the extent of modification-specific protein-binding interactomes investigated so far, homogenous, recombinant nucleosomal 12mer arrays carrying different chemical modifications, alone and in combination, were generated. These chromatin arrays were used as templates for affinity purification studies investigating protein-chromatin interactomics in the context of single modification and combinations thereof.

3.1.1 Native chemical ligation and alkylation of cysteine residues successfully introduce PTMs to recombinant *Xenopus laevis* histones

Two different approaches were used to introduce site-specific posttranslational modifications to recombinant *Xenopus laevis* histones. Modifications within the first 20 amino acids of the N-terminus of histone H3 and the first 22 amino acids of the N-terminus of histone H4 were introduced by native chemical ligation (NCL) (Shogren-Knaak and Peterson, 2004). The histone

methylation sites H3K9me1/-me2/-me3, H4K20me1/-me3 and H4R3me2 were successfully created by NCL. The products of the reactions were visualized by SDS-PAGE and mass spectrometry. This is exemplarily shown for the NCL of H4K20me3 in figure 3.1 A and B.

As a second approach to introduce PTMs, methyl lysine analogs (MLAs) were used. Here, lysine 27 of histone H3 was modified to H3K_C27me1/-me2/ and -me3. The method is based on the alkylation of an artificially introduced cysteine residue. The cysteine undergoes alkylation with an electrophilic ethylamine resulting in aminoethylcysteine, which is an analog of lysine (Simon et al., 2007). The completeness of the MLA reaction was verified by mass spectrometry using ESI-MS, exemplarily shown for H3K_C27me3 in figure 3.1 C.

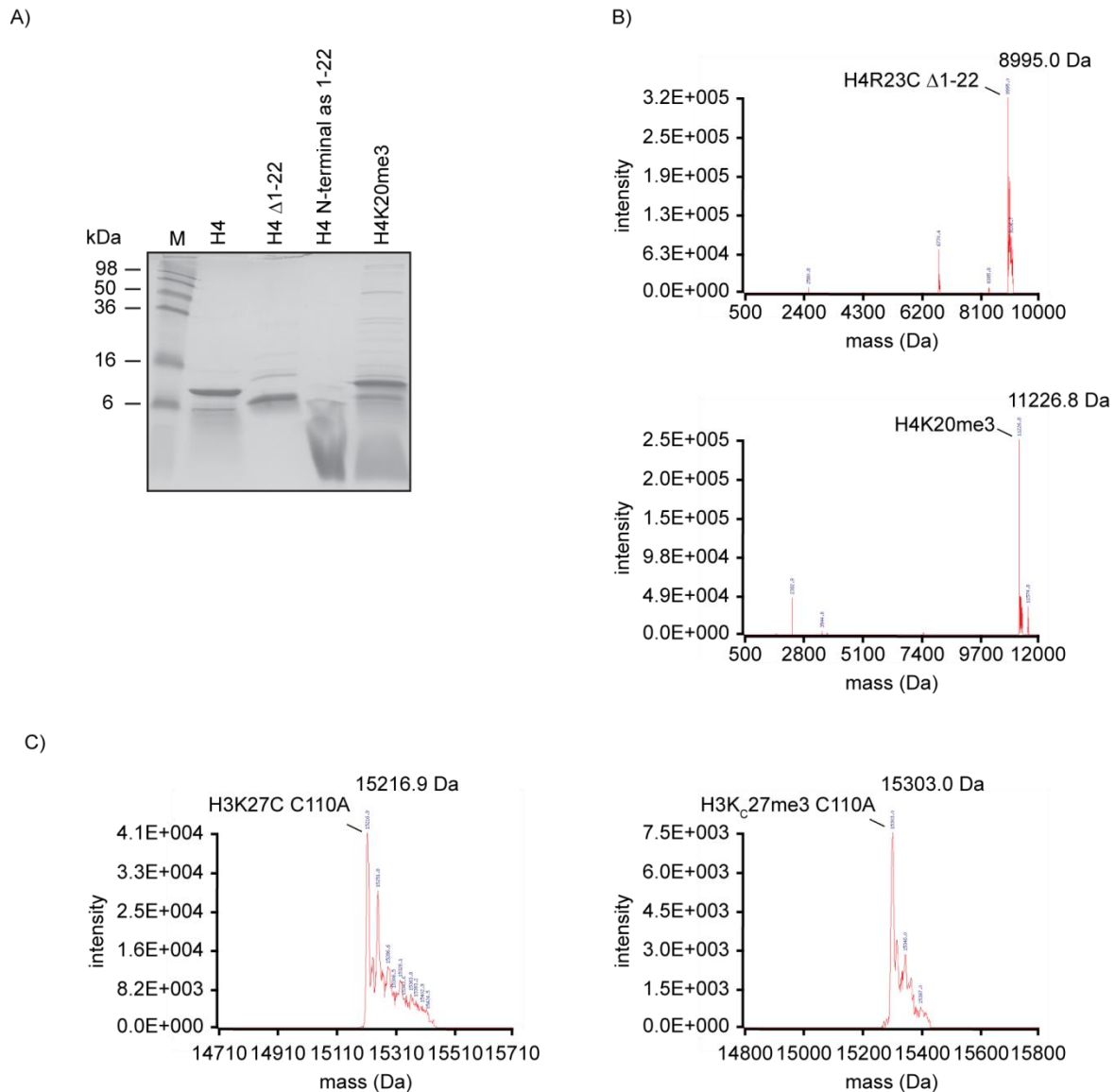


Figure 3.1 Introduction of posttranslational histone modifications.

A) SDS-PAGE analysis of the recombinant histone H4, H4R23C Δ 1-22, the thioester peptide composed of the N-terminal amino acids 1-22 of H4 and the ligated product H4K20me3. B) ESI-LTQ mass spectrometric analysis of H4R23C Δ 1-22 (upper part of the panel) and the ligation product H4K20me3

(lower part of the panel). C) ESI-LTQ mass spectrometric analysis of the non-alkylated H3K27C C110A (left part of the panel) and H3K₂₇me3 (right part of the panel).

3.1.2 Octamer-DNA molar ratios from 1.1 to 1.3 assemble fully saturated, reproducible nucleosomal arrays

The assembly of nucleosomal arrays is based on histone octamers that uniformly position on a DNA template. The optimal octamer-DNA molar ratio was determined for each chromatin species separately. An octamer-DNA molar ratio from 1.1 up to 1.3 was found to be sufficient to obtain a saturation of more than 90% of the DNA template with histone octamers. The success of the reconstitution was visualized by native agarose gel electrophoresis (Figure 3.2 A). The exact rate of saturation of the nucleosomal arrays was determined by a restriction enzyme digest using *Ava*I (Figure 3.2 B). *Ava*I cuts the DNA template of chromatin arrays within the 601 sequence only once per 200 bp repeat. Depending on octamer occupancy *Ava*I digestion results in mononucleosomes and free DNA.

We additionally tested the reproducibility of the chromatin reconstitution method using analytical ultracentrifugation. The distribution of sedimentation coefficients (S) of nucleosomal arrays, either fully or subsaturated (saturation level of 75%) were determined by sedimentation velocity and van Holde-Weischet analysis. As shown in figure 3.2 C, different fully saturated chromatin arrays displayed a similar sedimentation distribution, mainly in the range between 40 and 50 S. The similar sedimentation coefficient distribution of both replicates demonstrated that the reconstitution method generates highly reproducible chromatin arrays. In comparison, the subsaturated chromatin species showed a distribution of sedimentation coefficients in a lower range, between 30 and 40 S, indicating lower compaction of the chromatin species. The results verified that the saturation level of oligonucleosomal arrays can be reproducibly adjusted over the octamer-DNA molar ratio, generating arrays of comparable quality.

3.2 The statistical tool “significance A” determines significant protein fold enrichment cutoffs

To investigate the extent of changes in protein composition and binding to chromatin, affected by distinct chromatin modification patterns, chromatin-based affinity purification in combination with relative quantification by mass spectrometry was established in our lab (Figure 3.3) (Nikolov et al., 2011).

ChAP coupled to quantitative MS is based on the relative quantification of proteins binding to unmodified and modified chromatin arrays. This is possible by using SILAC-labeled HeLa nuclear extracts. For each modification state two biological replicates were performed in parallel

as a label swap experiment. In the “forward” experiment unmodified chromatin arrays were incubated with light nuclear extract and modified chromatin arrays with heavy isotope-labeled nuclear extract. For the “reverse” experiment modified and unmodified chromatin arrays were incubated with light and heavy nuclear extracts, respectively, thus obtaining inverse enrichment ratios.

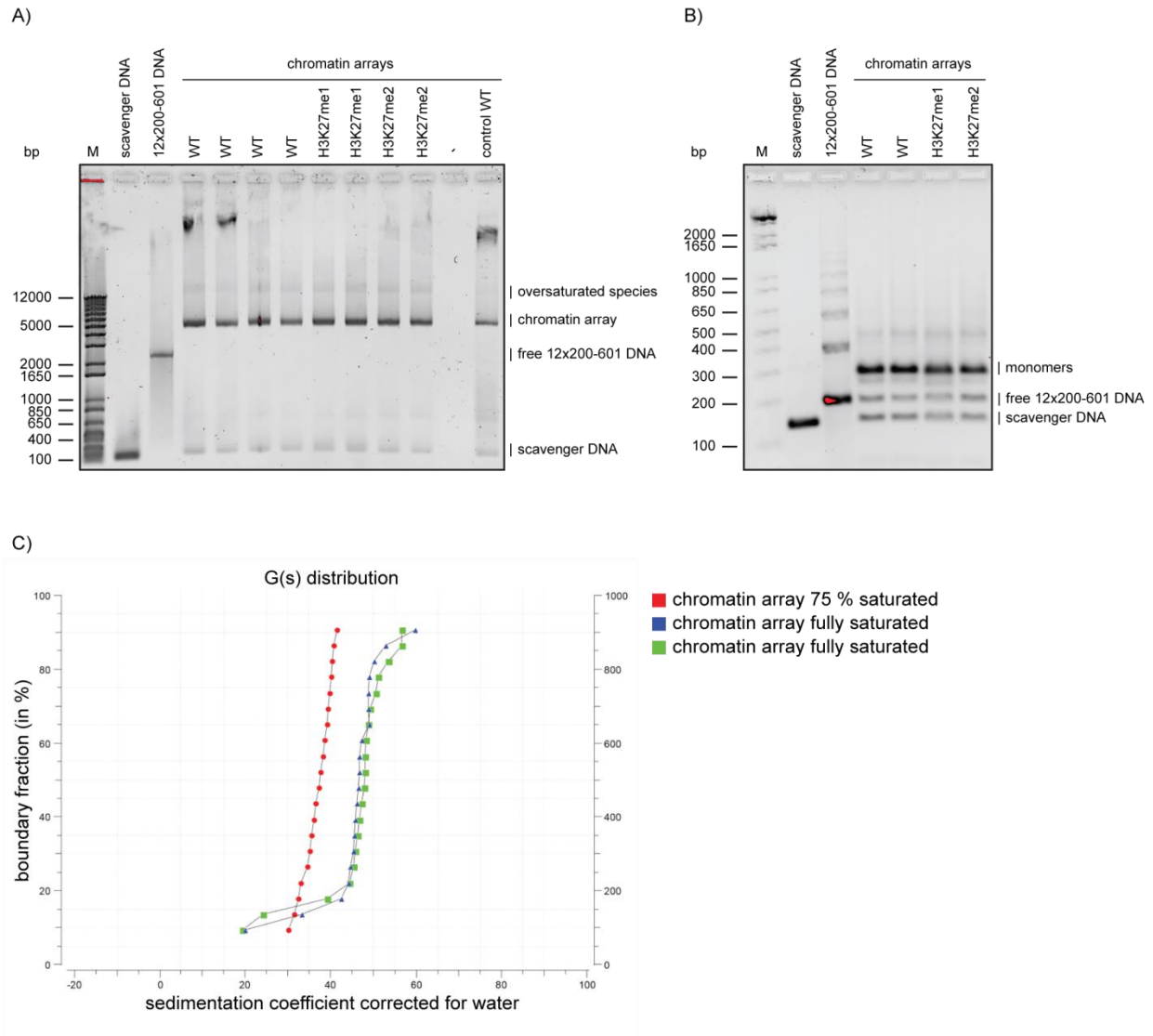


Figure 3.2 Quality controls of assembled chromatin arrays.

A) Native agarose gel electrophoresis of reconstituted chromatin arrays and free DNA. The samples were separated by gel electrophoresis on a 0.5% agarose gel using 0.2x TB buffer. The gel was stained with ethidium bromide. M defines the DNA size marker. B) Aval digestion of reconstituted chromatin arrays and free DNA. Digested samples were separated by agarose gel electrophoresis on a 2% agarose gel using TBE buffer. The gel was stained with ethidium bromide. C) Analytical ultracentrifugation of fully saturated and 75% saturated chromatin arrays. The distribution of sedimentation coefficients was obtained using sedimentation velocity and van Holde-Weischet analysis (analyzed by Dr. Ron Finn, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany).

The method enables the identification of all proteins binding to chromatin and that are specifically regulated by the underlying chromatin modification. Throughout this study, I will refer to this pool of proteins as the “chromatin-binding interactome”. By investigating different modification patterns several interactomes were generated by performing experiments with individual chromatin modifications. Hereby, a statistical evaluation strategy ensuring a valid interpretation of the results not only for individual experiments but also for comparison of different interactomes is of immense importance.

In previous studies, a fixed fold enrichment cutoff was defined and applied to all individual experiments, not meeting the requirements of a significant determination of a fold enrichment cutoff for each individual experiment. To verify significant protein enrichment, I optimized the strategy of statistical evaluation to faithfully investigate and interpret the interactomes brought about by twelve different chromatin modification patterns.

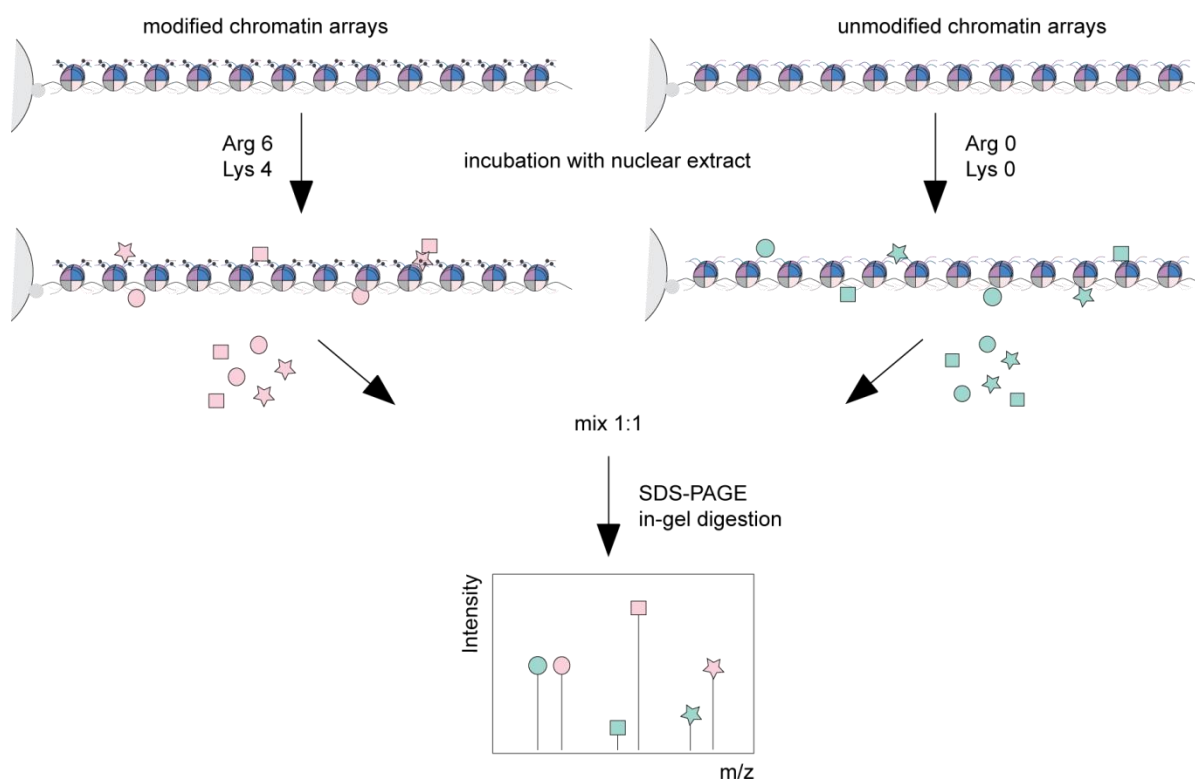


Figure 3.3 Workflow of SILAC-based chromatin affinity purification coupled with mass spectrometry.

Unmodified and modified chromatin arrays are incubated with light and heavy SILAC-labeled nuclear extracts from HeLa cells, respectively. Chromatin arrays and their associated proteins of both pull-downs are eluted, mixed in a 1:1 ratio, treated with benzonase and finally separated by SDS-PAGE. Proteins are in-gel digested using trypsin. The resulting peptides are analyzed by LC-MS/MS. Identification and quantification is achieved using the MaxQuant software.

To determine the impact of histone PTMs on the chromatin-binding interactome, the proteins associated with modified and unmodified chromatin arrays were compared and their relative enrichment level was determined. Quantification using SILAC-labeled nuclear extracts assumes that unspecific binders, i.e. proteins binding to both templates to the same extent, feature a heavy/light ratio (H/L ratio) value of 1. However, due to small variations inherent to all biological preparations this value might vary from one experiment to another. Considering all performed experiments this theoretical heavy/light ratio of 1 ($\log_2 1 = 0$) for unspecific binders in fact ranged from $-3.2 \log_2$ up to $2.3 \log_2$ (visualized in chapter 3.3 for each individual experiment) as calculated by applying a statistical significance test as described in the following.

The definition of a fixed fold enrichment cutoff is not sufficient to faithfully reflect which proteins are significantly enriched or depleted when two different data sets are compared. To overcome this problem, the statistical significance of individual quantification results was defined using the Perseus' "significance A" algorithm (Cox and Mann, 2008). This algorithm determines the values that are significant outliers relative to a certain population. The "significance A" algorithm was used by applying a two-sided False Discovery Rate estimation to a p-value of 0.1 and a Benjamini-Hochberg correction.

To determine significant enrichment values, SILAC ratios of the forward and the reverse experiment were averaged. This resulted in a narrowed distribution range of H/L ratios of the protein population bound to chromatin. This can be visualized when comparing the distribution of H/L ratios of proteins resulting from the individual experiments with the distribution of the averaged forward and reverse experiments (Figure 3.4 A-C). Consequently, an increased number of significant outliers was calculated compared to calculations using non-averaged SILAC ratios, as significant values are determined as outliers relative to the spread of the input population (Figure 3.4 A-C).

In order to visualize the results, SILAC ratios of forward and reverse experiments were scatter plotted on the x- and y- axis, respectively (Figure 3.4 D). The advantage of this presentation style is the complexity of information that can be visualized.

First, the protein-binding affinity to modified chromatin arrays is demonstrated. Therefore, H/L ratios of the reverse experiment were inversed due to the label-swap approach. Data were \log_2 scaled, where positive values of the H/L ratio represent an increased protein-binding affinity, whereas negative H/L ratios represent a lower binding affinity to modified chromatin arrays compared to the unmodified chromatin species. Thus, the scatter plots display proteins whose binding is either increased or decreased as a consequence of a specific histone modification state.

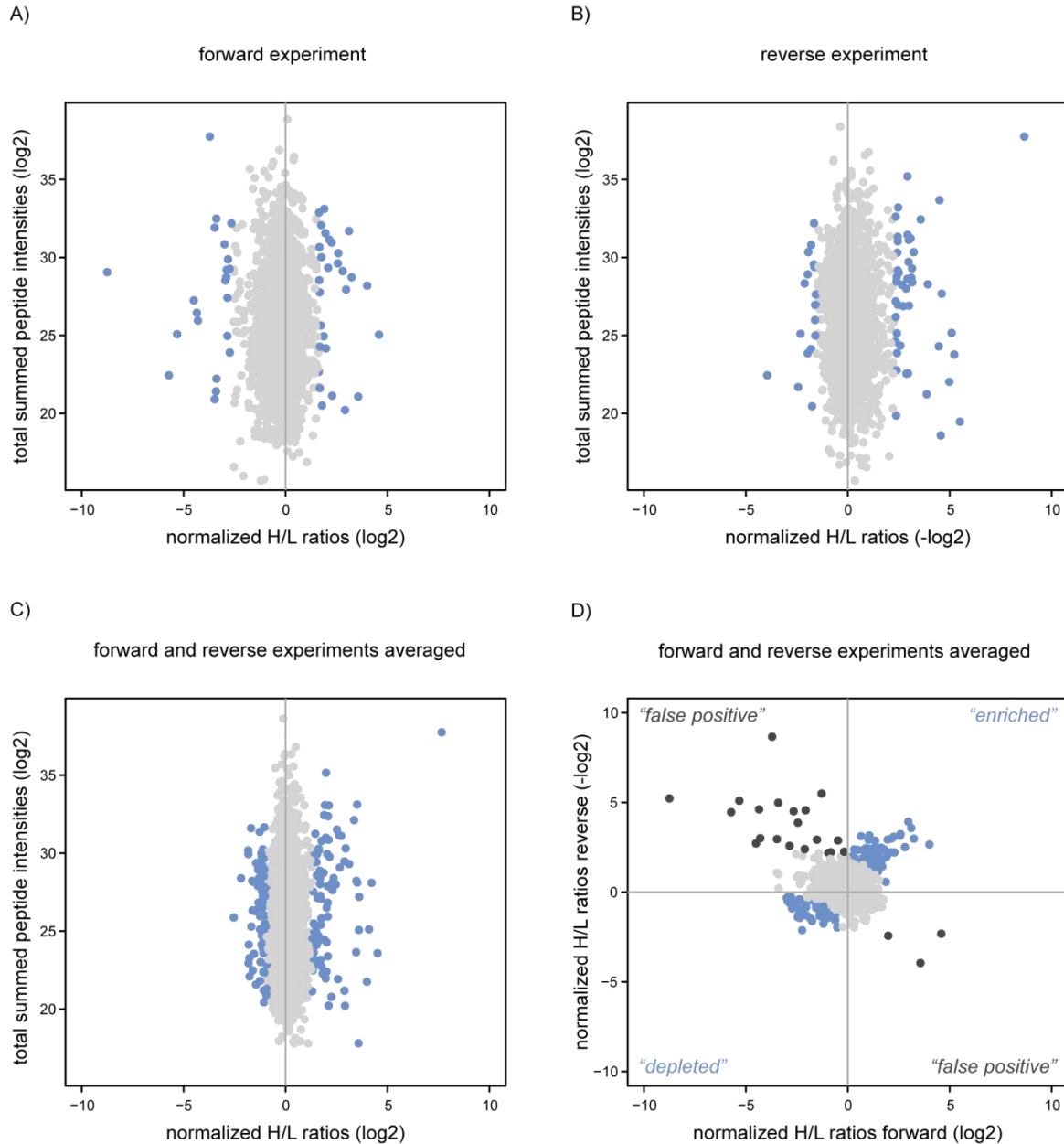


Figure 3.4 Determination and visualization of significantly regulated proteins using the example of the H3K9me3 interactome.

A) Log₂ SILAC ratios of proteins identified in the forward experiment are plotted on the x-axis. Total summed peptide intensities of the forward experiment are plotted on the y-axis in log₂ scale. Proteins identified as significant outliers by application of the "significance A" algorithm are marked in blue. B) The same as in A) but with values of the reverse experiment. SILAC ratios are inverted. C) The same as in A) but with averaged values of the forward and the reverse experiments. D) Log₂ scaled H/L ratios of proteins of the forward experiment are plotted on the x-axis and of the reverse experiment on the y-axis. Statistical analyses were performed with averaged H/L ratios of the datasets. Proteins determined as significantly enriched are displayed in the upper right square, while proteins significantly depleted are displayed in the lower left square of the plot. Proteins determined as false positives are displayed in the upper left and the lower right square. Significantly recruited and depleted proteins are colored in blue. False positive proteins are marked in black. Proteins that were significantly regulated but only identified in either the forward or the reverse experiment are not plotted.

Second, the scatter plots visualize the correlation of both biological replicates. In case of positive correlation, for both replicates proteins displaying the same trend in terms of binding affinity to modified chromatin arrays and are located in the upper right (increased binding) and in the lower left square of the scatter plot (decreased binding). Negative correlation is observed for proteins displaying an opposed H/L ratio in both biological replicates. Those proteins are located in the upper left and the lower right square of the scatter plot (Figure 3.4 D). Accordingly, negative correlation is indicative of false positives. Therefore, significant outliers that displayed a diametrically opposed significance score in the forward and reverse experiments were eliminated from further analysis (Figure 3.4 D).

Third, statistically evaluated information is represented by a color code. Significant outliers that correspond to proteins significantly regulated by a certain modification state are marked in blue, whereas proteins constituting the background are labeled in grey. Proteins identified as false positive are colored in black.

With this statistical evaluation strategy all further introduced chromatin-binding interactomes were analyzed and the results were visualized as described.

3.3 ChAP coupled to quantitative MS enables the investigation of chromatin-protein interactomics and provides insights into chromatin modification crosstalk

The first steps towards investigating whether, and even more importantly, how posttranslational chromatin modifications affect the structure of chromatin and downstream biological processes, such as transcription, replication and cell division, is the investigation of the impact modifications have on the protein environment of the chromatin loci they mark. Heterochromatin is characterized by a tightly packed form of chromatin and thought to be mainly transcriptional silenced. However, it is specifically marked by a variety of posttranslational modifications and yet, it is not fully understood how and whether these modifications contribute solely to the molecular mechanisms of establishment and maintenance of the heterochromatic state. To investigate the extend of biological functions correlating with heterochromatin I focused on chemical modifications in literature so far associated with this chromatin stage. The following chapter introduces the protein-binding interactomes mapped in the context of different histone methylations, DNA methylation and combinations thereof.

3.3.1 Individual chromatin modification states recruit specific chromatin-binding interactomes

Interactomes associated with 10 different chromatin modifications were characterized. For each modification state between 1,117 and 2,785 chromatin-bound proteins, including their isoforms, were identified. All analyses together identified more than 6,000 proteins. Of these, close to 500 were found to be regulated by the chromatin modifications. Each modification state impacted the binding properties of a specific set of factors. Moreover, each dataset included several proteins that were not known before to associate with specific histone modifications. A comprehensive list of factors whose binding to chromatin was impacted by the presence of a certain modification is provided in table 3.1. These data shed light on how extensively the chromatin interactomes are regulated by histone posttranslational modifications. The individual protein-binding interactomes for each histone modification state will be discussed in the following sections.

H3K9me1 interactome

Genome wide high resolution maps revealed that H3K9me1 is mainly associated with transcriptional active gene loci (Barski et al., 2007; Wang et al., 2008). The predominant absence of H3K9me1 at constitutive heterochromatin has been demonstrated before by using chromatin immunoprecipitation (ChIP) and antibody based detection methods in ES cells (Peters et al., 2003). In contrast, antibody based staining of MEF cells demonstrated that H3K9me1 is located at pericentric heterochromatic regions (Sims et al., 2006). To my knowledge, proteomic studies investigating protein interactions to H3K9me1 on a global scale have not been performed.

In this proteomic study, H3K9me1-modified chromatin arrays recruited six and repelled seven factors from binding (Figure 3.5 A, Table 3.1). Although H3K9 mono-methylation was mainly shown to be linked to gene activation, several factors connected to heterochromatin were identified to be specifically recruited by H3K9me1. Among them were the heterochromatin protein HP1 α (CBX5) and UHRF1. Both are known to be involved in heterochromatin establishment and maintenance and have been shown to bind preferentially to K9me3 of histone H3 (Lachner et al., 2001; Nady et al., 2011). The splicing co-activator SRRM1 (Blencowe et al., 1998) was also recruited to H3K9me1-modified chromatin, indicating that not only heterochromatin associated factors were recruited.

To my knowledge UHRF1 is the only factor of the interactome that has been described so far to directly interact with H3K9me1 (Nady et al., 2011). Protein interaction analysis using the database STRING (Jensen et al., 2009) predicted no protein-protein interactions besides the interaction between CBX5 and PRMT5 (Figure 3.5 B, upper panel). Thus, it remains unclear, whether the six proteins bind as single factors or form a multi subunit complex.

In contrast, the transcription factors NFIB, NFIC, NFIX and E4F1 were significantly repelled from binding, as was INO80E, a component of the INO80 chromatin remodeling complex. The protein-protein interactions between NFIB, NFIC and NFIX and between INO80E and MCRS1 were predicted by STRING analysis (Figure 3.5 B, lower panel), suggesting that H3K9me1 regulated factors are biologically linked, which verifies the specificity of the approach.

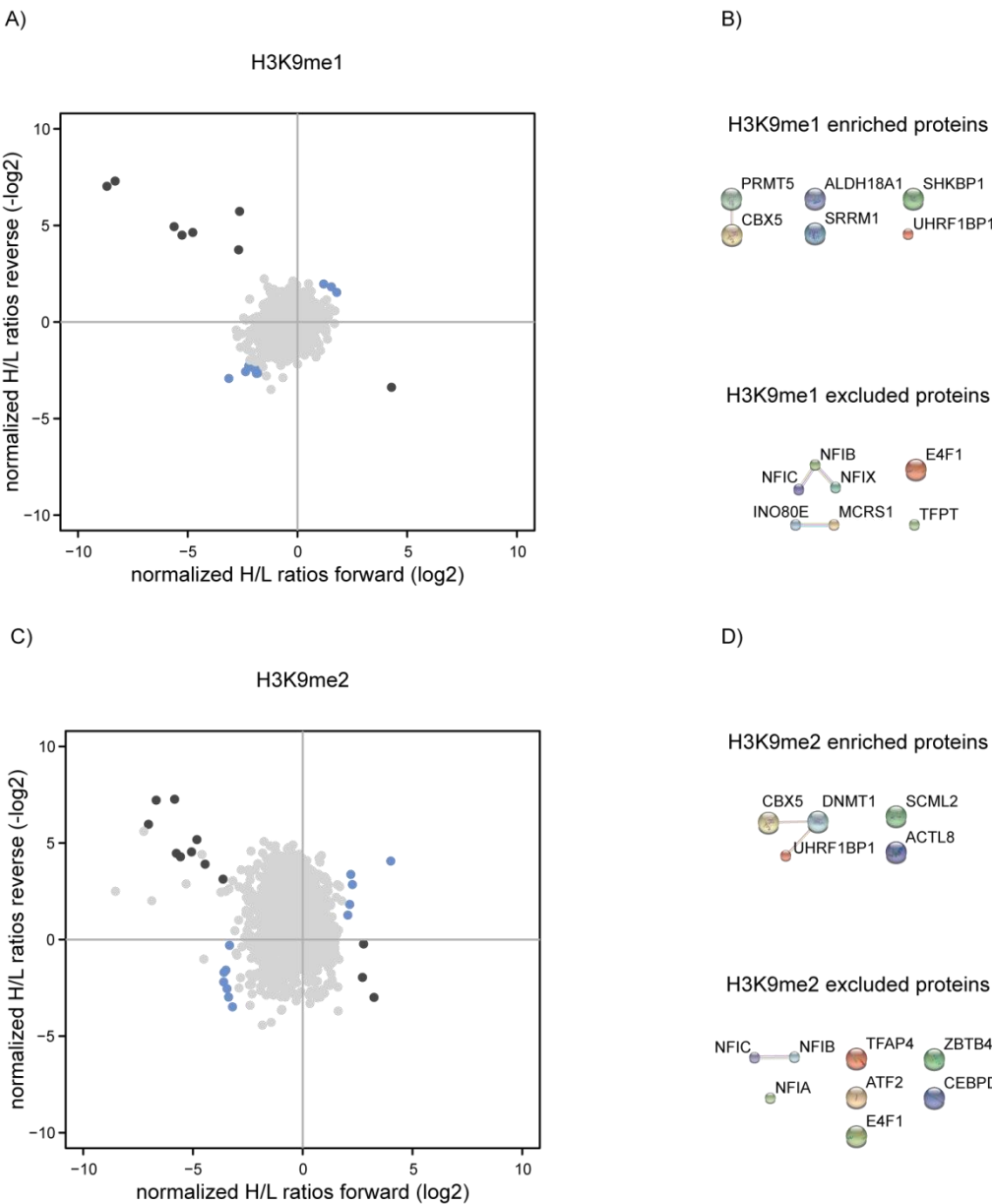


Figure 3.5 Protein-binding interactomes of H3K9me1 and H3K9me2 chromatin.

A) H3K9me1 interactome. Proteins are plotted by their log₂ SILAC ratios of the forward experiment on the x-axis and the reverse experiment on the y-axis. Significantly recruited and depleted proteins are colored in blue. False positive proteins are marked in black. B) STRING protein-protein interaction networks of factors significantly recruited by H3K9me1 (upper part of the panel) and excluded proteins (lower part of the panel). C) H3K9me2 interactome. Data are represented as described in A. D) Protein-protein interaction networks from STRING analysis of proteins significantly affected by H3K9me2. Recruited

proteins are shown in the upper part of the panel, whereas excluded proteins are represented in the lower part of the panel.

H3K9me2

The chromatin-binding interactome of H3K9me2 introduced here is represented by the results of the forward experiment only, as data evaluation revealed a non-normal distribution of SILAC ratios of chromatin-bound proteins identified in the reverse experiment.

Chromatin di-methylated at lysine 9 of histone H3 recruited five and depleted eight factors from binding (Figure 3.5 C, Table 3.1). As the modification is known to be linked to gene repression (Wang et al., 2008), it is not surprising that nearly all factors recruited have been shown to be involved in maintenance of a repressive transcriptional state. These proteins comprise CBX5, UHRF1, DNMT1 and the putative polycomb group protein SCML2 (Montini et al., 1999). HP1 proteins, such as CBX5 have been shown to bind to H3K9me2 N-terminal peptides (Jacobs and Khorasanizadeh, 2002; Lachner et al., 2001). Also, UHRF1 has been demonstrated to bind to H3K9me2/-me3 (Liu et al., 2013). Moreover, UHRF1 was identified to interact with DNMT1, (Bostick et al., 2007; Sharif et al., 2007), suggesting a role for UHRF1 in DNMT1 recruitment to H3K9me2-marked chromatin regions. A link between CBX5, UHRF1 and DNMT1 was further suggested by STRING analysis (Figure 3.5 D, upper panel). There is one published study that screened for H3K9me2-binding proteins using N-terminal modified peptides of H3 (Chan et al., 2009). Comparing the findings of this study to mine, CBX5 was found to be the only overlapping factor identified to bind in the context of H3K9me2.

The transcription factors CEBPD, TFAP4, ATF2 and the zinc finger protein ZBTB43 were depleted from binding to H3K9me2 in addition to the transcription factors also excluded from binding to H3K9me1-modified chromatin arrays. STRING analysis did not reveal any association of these factors (Figure 3.5 D, lower panel).

H3K9me3

H3K9me3 is the most intensively investigated repressive histone modification. It marks pericentric heterochromatin (Barski et al., 2007; Fischle et al., 2003a; Lachner et al., 2003; Wang et al., 2008). In my study, no other modification state affected as many proteins as trimethylated lysine 9 of histone H3. Altogether, 59 proteins were specifically recruited and 72 repelled from binding (Figure 3.6 A, Table 3.1). Previous studies using different methodologies, already identify a set of H3K9me3-binding factors comprising among others the different isoforms of HP1, CBX1, CBX3 and CBX5 (Lachner et al., 2001), ADNP (Mosch et al., 2011) as well as UHRF1 (Karagianni et al., 2008) and MPHOSPH8 (Kokura et al., 2010). Besides single

Table 3.1 Proteins recruited and excluded from binding to chemically modified chromatin.

Proteins are listed according to their enrichment ratios (H/L ratios) starting with the highest value on top. Proteins significantly recruited by a certain modification state are written in black, proteins excluded are written in blue. Grey marked proteins indicate that the proteins were already found to interact with the respective modification in one of the following studies (Bartke et al., 2010; Bluhm et al., 2016; Bostick et al., 2007; Engelen et al., 2015; Kunowska et al., 2015; Nady et al., 2011; Nikolov et al., 2011; Oda et al., 2010; Vermeulen et al., 2010)

H3K9me1	H3K9me2	H3K9me3	H3K27me1	H3K27me2	H3K27me3	H3 Δ1-20	H4K20me1	H4K20me3	H4R3me2	meCpG	H3K9me3- H4K20me3	H3K9me3- meCpG
SHKBP1	UHRF1	CBX1	GCC2	JADE2	EIF2B5	FAM114A1	SCML2	DNAH8	POLR3H	UHRF1	DNAH8	DSC3
SRRM1	CBX5	UHRF2	CNOT2	KIAA1524	PHF1	CAPN1	LCN1	ORC2	POLR3D	DNMT1	LRWD1	CALML5
ALDH18A1	ACTL8	CBX3	BAG6	LGALS7	FGF2	TALDO1	HDGFRP2	LRWD1	SNC73	SCML2	CBX1	TGM3
PRMT5	DNMT1	CBX5	RSF1	PHF1	FRMPD3	MIF	ZRANB2	ORC3	POLR3G	RALGAP2	CBX5	CSTA
UHRF1	SCML2	USP3	DYM	GBP1	LRWD1	ZGPAT	PC4	HMGNS	POLR1C	ZHX2	ORC3	CDK2AP1
CBX5	NFIC	SPIN2B	DFNA5	SLC25A3	ORC2	HDGF	RMND1	ARG1	E2F6	ZBTB33	CHAF1B	DNCL1
MCRS1	CEBPD	POGZ	PRIC295	MAGO	ORC3	CAST	SREK1	ORC5	NFIC	USP7	ORC2	TRAM1L1
NFIB	NFIB	CHAF1A	CNOT1	SHPRH	ORC5	PSAT1	USP7	SIRT6	POLR3E	ACTL8	POGZ	
NFIX	ZBTB43	CHAMP1	IPO11	PFN1	ORC3L	WARS	ENO1	EIF5	POLR2E	ZHX1	CBX3	
E4F1	NFIA	CHAF1B	KIF11	PRPF40A	CBX8	SYAP1	RSRC2	APEX1	ACTL8	ZHX3	ORC5	
INO80E	TFAP4	NIPBL	CNOT11	SRSF6	PRC1	PARK7	CIR	ORC3L	USP7	UHRF2	SPIN2B	
TFPT	E4F1	SEPT7	XPO1	PGK1	ABCF1	FMNL1	NAP1L4	SYAP1	NOLC1	SEMGI	NIPBL	
NFIC	ATF2	SEPT9	FBXL6	SAP18	EIF2S2	ANXA5	RPRC1	PHF1	NAA40	CUEDC1	CHAF1A	
		SIRT6	UBA5	EIF4A3	TOPBP1	POU2F1	PSMD3	RNF213	LMNA	MBD2	UHRF2	
		SEPT2	SIRT6	ACIN1	WDR5B	TPT1	PHF8	EIF5B	NCL	ZBTB12	SPIN1	
		SPIN1	RAVER1	PHF21A	TCOF1	BCLAF1	INO80C	USP3	CHD1	MTA2	CHAMP1	
		ACTL8	GET4	CLIC1	EIF2S3	NBN	MYC	EIF2S2	SCML2	UBE2D2	ORC3L	
		DNMT1	TTI1	ZC3H11A	PRC1	PNP	MLXIPL	EMG1	ZCCHC10	GATAD2B	ADNP	
		ADNP2	CNOT7	RNPS1	EIF2S1	CKB	POLG	EIF2S1	TADA1	CDK2AP1	MAU2	
		LRWD1	ZKSCAN3	TKT	MAFK	ALDOC	TKT	ORC6	RAD51AP1	HYDIN	NOLC1	
		USP7	SPTY2D1	PAXIP1	PBX2	GD12	PGK1	TXLNA		ZNF687	JADE2	
		HMGNS	PSMA7	CHAF1B	ELF2	AHCY	PNISR	GOLT1B		FIZ1	ING5	
		ORC3L	POLR3K	ISL2	KIF2A	EIF4H	PHF21A	EIF2S3		ZMYND8	PHF5A	
		ORC2	PRC1	E4F1	CDYL2	CXorf38	ZNF580	KIF2A		RBBP7	SF3B1	
		TCHH	VWA9	AEBP2	AAR2	ZNF451	RFXANK	RAD51AP1		CTSD	SHPRH	
		ORC3	CHD1	INO80E	PLK1	MAGO	HES1	KIF2C		GATAD2A	UBTF	
		UHRF1	TAF11	TFPT	RBM7	FAM48A	INO80D	ABCF1		CHD3	SF3A1	
		RAD51AP1	RECQL	SP3	XPO5	VRK3	DHX36	BCR/ABL f.		RBBP7	SF3B4	
		ADNP	MED10	NR2C1	KIF5B-ALK	DLG3	NCAPD2	TXLNG		PGBD3	PMVK	
		ZHX1	HSP90AB1	HEATR2	EXOC2	FAM134C	NCOR2	KIFC1		MTA3	FRMPD3	
		EIF5	TP1	ZMYM1	RPLP2	NACA	NEIL2	ZNF598		BLMH	SF3A3	
		ORC5	CENPF	SP1	NOLC1	TMPO	CHTOP	PDIA5		CHD4	SEPT10	
		MIER2	SETX	ZNF629	SPATA5L1	CEBPD	ZC3H14	NFIX		RBM4	SF3B2	
		EMG1	CKAP2	NFIB	OXSRI	GOLGB1	SNRPD3	C19orf47		NCOA6	SEPT9	
		SEPT6	PFN1	ZNF770	NPM1	FMNL1	SRSF6	NFIA		ZBTB4	PGK1	
		MPHOSPH8	MED16	CD3EAP	RPRC1	SARS	SAP18	POLR1B		POGZ	SEPT6	
		ATRX	GF3C4	POLR1E	PCMT1	ATL3	PNN	PKP2		RBBP4	C9orf78	
		MAU2	HSPA9	KMT2C	JADE2	LMNA	DDX21	ANKRD12		MECP2	LSM2	
		KIFC1	GTF3C5	PRDM10	YKT6	ARPC2	MAGO	BRCA1		MTA1	USP3	
		SMCHD1	LDHB	POLR1A	EPPK1	ALDH18A1	RPL7L1	E4F1		SEMGI	EIF4G1	
		SEPT8	ZBTB1	PARD3	TINF2	UIMC1	SF3B4	ANKRD32		BACH1	EIF5	
		CDYL	MARK2	POLR1B	NAP1L1	CBR1	SNRPB	BARD1		CHAMP1	ING4	
		SCML2	PRDX1	ESRRA	UBL4A	ASNS	PURB	AMY1A		WHSC1	RAB6A	
		KIF2A	LDHA		SYDE1	CPNE1	CGGBP1			POLR3F	BRCA1	
		ABCF1	ACTN4		SCYL2	CUL1	MAP4			E2F2	SEPT7	
		EIF2S2	PKM		HTRA1	TADA3	ZNF354A			SIN3A	CHD1L	
		MORC2	TKT		MAP2K2	ARHGEF3	SNRPA1			CLOCK	JADE3	
		EIF2S1	EEF2		CHERP	BPTF	ZC2HC1A			POLR1C	RAD51AP1	
		EIF2S3	PRDX6		CBFB	ARPC4	SF3A3			BHLHB2	BARD1	
		MIER1	PNP		UBTF	PES1	ZBED6			RPRD1B	BABAM1	
		ELF1	EEF1B2		PCSK9	POM121C	PURA			ZNF131	SEPT2	
		CDYL2	PGAM1		C11orf57	LIG4	SF3B2			BRMS1L	BRCC3	
		ZC3HAV1	PSMA6		PHF8	SHCBP1	SF3B1			MITF	ZBED1	
		JMJD1C	ENO1		ACACA	MTF2	ANKRD12			ELF1	SDHB	
		LRIF1	G2E3		GTF2E2	POLD3	SF3A1			SAP30	ACACA	
		WHSC1L1	PPIA		TYMS	PUM1	RSL1D1			SUPT3H	BRE	
		ZNF280D	IGHG4		ZHX2	RIOK3	ATRIP			SAP130	CSTA	
		CDKN1A	ALDOA		MBTPS1	RSL1D1	ACACA			ARNTL	ASF1B	
		ZMYM2	GNB2L1		AEBP2	SMARCA5	BRE			SUDS3		
		RNF213	DSC2		JARID2	CCDC59	BRCA1			CGGBP1		
		ETV6	ANXA1		G2E3	MRPL13	BARD1			ELF2		
		TINF2	SERPINB3			PUM1	ANKRD11			TAF1D		
		INO80D	MMS22L			COBL				ZSCAN20		
		PCNA	TGM3			MOB4				MAX		
		CACYBP	SFN			NFATC2IP				ARID4A		
		TAF1	CDSN			FBXW11				BRMS1		
		TAF5	EPPK1			POP4				AP2A1		
		TAF6	POF1B			PLEKHA4				MLXIPL		
		WDR77	TADA2B			CENPV				ING2		
		CEBPG	DSC3			HMG3B3				E2F1		
		ZNF644	TF			HMG20A				ING1		

H3K9me1	H3K9me2	H3K9me3	H3K27me1	H3K27me2	H3K27me3	H3 Δ1-20	H4K20me1	H4K20me3	H4R3me2	meCpG	H3K9me3- H4K20me3	H3K9me3- meCpG
		PRMT5	YOD1			DDX5-ETV4 f.				MNT		
		RFX1	CAT			SIN3B				S100A11		
		PAXIP1	GSTP1			TIMP2				HES1		
		WIZ	LGALS7			GNBP2				MLX		
		TAF1A	IGHG1			FNBP4				RAD1		
		POLR3E	PRDX2			CLASP2				USF1		
		KMT2D	FABP5			DDX52				USF2		
		POLR3D	CSTA			USP7				ZBTB14		
		ZBTB44	IGKC			IFI16				ATP6V1F		
		ZBTB39	ARG1			FOXN2				ZBTB1		
		POLR3GL	DSC1			CLASP1				AP2B1		
		UCHL5	SPRR1A			SUZ12				SYN1		
		INO80	SPRR1B			KAT7				EPPK1		
		ZNF444	IGLC1			TARS				SYT1		
		NCOA6	TGM1			EZH2				ATP6V0D1		
		ACTR8	S100A14			EED				SFN		
		INO80C				USP1				ATP6V1B2		
		POLR3A				KMT2C				LGALS7		
		TFPT				DNMT1				SV2A		
		PATZ1				MEAF6						
		ACTR5				WDR5						
		INO80E				DDIT4						
		CRCP				CPNE8						
		NFRKB				ACTL8						
		INO80B				C10orf12						
		TBRG1				BRD2						
		SP3				TWTF2						
		EHMT2				RAI14						
		BAZ2A				NS5ATP4A						
		EHMT1				PHF14						
		POLR3C				CCDC138						
		POLR3B				JADE3						
		TIAL1				SVIL						
		POLR3K				LRRFIP2						
		CEBPD				CAPRIN1						
		POLR3F				TOB1						
		HIVEP2				PRKD2						
		POLR3G				IGF2						
		ZNF148				MYL6B						
		CEBPB				ACACA						
		POLR1D				UHRF1						
		POLR3H				BAZ1B						
		POLR2H				KRI1						
		POLR2E				CDSN						
		KLF16				IGKC						
		POLR1C				IGF1						
		ZNF282				DEFA3						
		POLR2K										
		POLR2L										
		POLR1A										
		CYR61										
		POLR1B										
		NFIX										
		ZMYM1										
		NFIC										
		TAF1B										
		FAM129B										
		PHF20L1										
		NFIB										
		NFIA										
		E4F1										

factors, protein complexes were also recruited by H3K9me3. In agreement with previous findings (Bartke et al., 2010; Vermeulen et al., 2010), several members of the human ORC complex were found in the dataset. Additionally, a set of factors not shown to bind to H3K9me3-modified chromatin before was identified. For example, four out of the five subunits of the septin complex were found. Protein-protein interaction analysis using the STRING database predicted an indirect connection of SEPT2 to CBX5 via ADNP (Figure 3.6 B). Moreover, this analysis predicted a network including most of the proteins specifically recruited by H3K9me3, thus validating the approach in terms of specificity and known biological context. Furthermore, the STRING analysis indicated recruitment of factors completely new in the context of H3K9me3

binding and therefore argued for a gain in knowledge using chromatin arrays over e.g. histone peptides for identification of histone PTM-binding proteins.

Proteins depleted from binding seemed to be organized in complexes as well. For example, seven subunits of the INO80 complex and several members of the BRAC1-core RNA polymerase II complex were prevented from binding. These findings appeared to be directly connected, as 57 of the 72 factors excluded from binding to H3K9me3 are included in an extensive protein-protein binding network based on STRING analysis (Figure 3.6 C).

H3K₂₇me1

The mono-methylation of lysine 27 of histone H3 has been shown to have a higher prevalence at active promoters compared to silent ones (Barski et al., 2007). Other studies mapped the modification to be significantly present in heterochromatin (Jacob et al., 2010). Chromatin affinity purification using H3K₂₇me1 modified chromatin revealed 19 factors significantly recruited and 68 excluded from binding (Figure 3.7 A, Table 3.1). Among them, four members of the CCR4-NOT transcription complex, which is a key regulator of gene expression (Collart, 2016), were found (Figure 3.7 C). Other proteins identified are also involved in positive regulation of gene expression. These include the chromatin assembly factor RFS1 (LeRoy et al., 1998; Shamay et al., 2002) and GCN1L1 (Marton et al., 1997). Also recruited were the deacetylase SIRT6, the ribonucleoprotein RAVR1, which acts as a splicing co-repressor and thereby modulates alternative splicing events (Gromak et al., 2003; Plafker and Macara, 2000), and the proteins XPO1 and IPO11 that function in nuclear export and import, respectively (Kudo et al., 1997). Surprisingly, three of the identified factors, DYM, GET4, and GCC2, are involved in processes connected to the Golgi apparatus and are known to locate to the cytoplasm (Dimitrov et al., 2009; Mariappan et al., 2010; Reddy et al., 2006).

Factors excluded from binding to H3K₂₇ mono-methylated chromatin feature a broad range of enzymatic activities and thus functional diversity. For instance, there were proteins depleted from binding that promote transcription such as GTF3C5, GTF3C4 or POLR3K and proteins connected to RNA polymerase II activity such as TAF11 and MED10. Many factors are connected to cytoskeleton associated functions i.e. PFN1, CKAP2, PRC1 and CENPF. Enzymatically active proteins included for example the pyruvate kinase PKM, the hydrolase YOD1 that can remove conjugated ubiquitin from proteins, and the G2/M-phase specific E3 ubiquitin protein ligase G2E3. Two members of the 20S proteasome were identified as well. Even though there was a high diversity of expelled factors, protein-protein interactions predicted by STRING indicated that 34 of them are connected to one another (Figure 3.7 B).

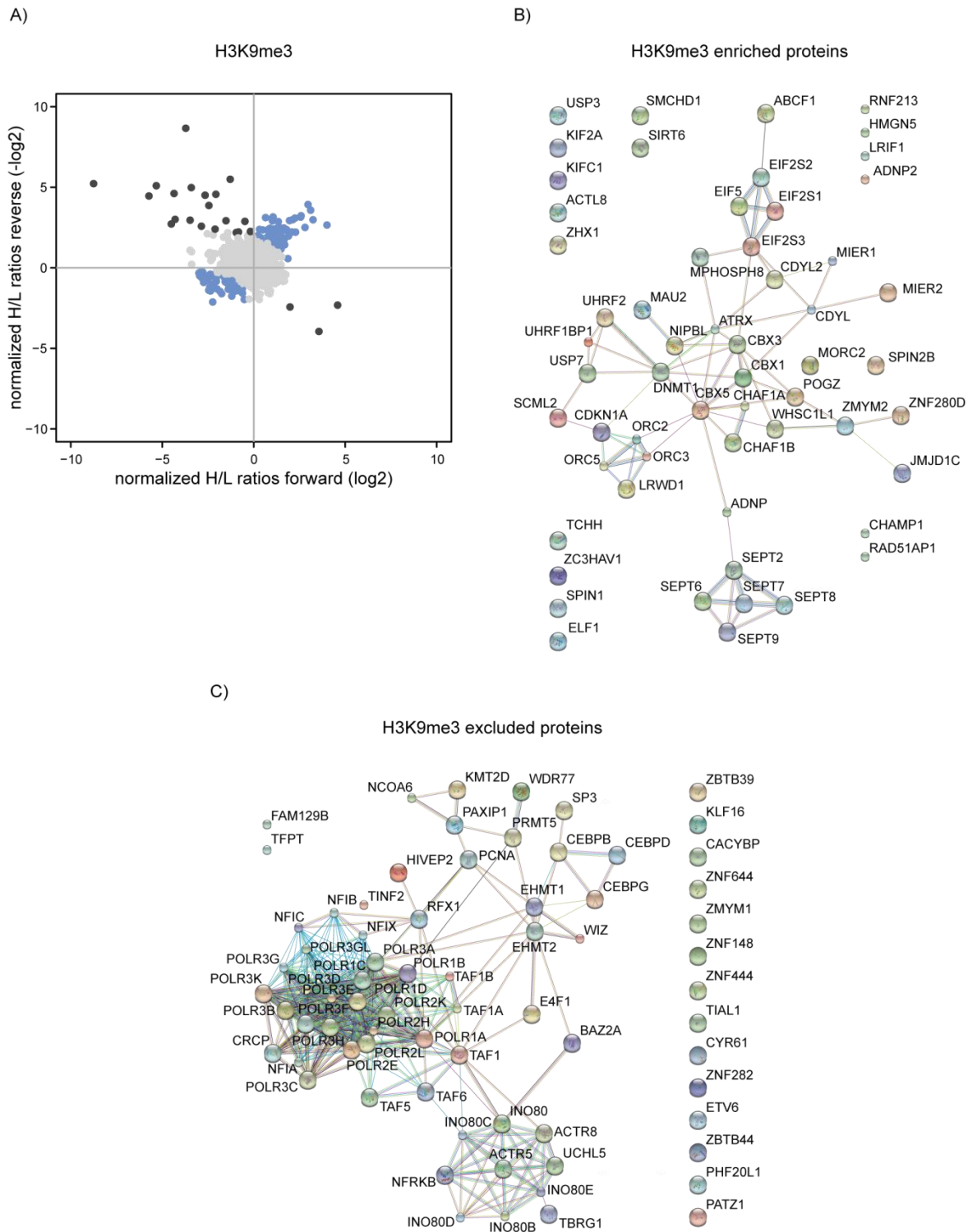


Figure 3.6 Protein-binding interactome of H3K9me3-modified chromatin.

A) H3K9me3 interactome. Proteins are plotted by their \log_2 SILAC ratios of the forward experiment on the x-axis and the reverse experiment on the y-axis. Proteins significantly recruited and excluded from binding to chromatin are colored in blue. False positive proteins are marked in black. B) STRING protein interactome of proteins significantly recruited by H3K9me3. C) STRING protein-protein interactions of factors significantly excluded by H3K9me3.

H3K_C27me₂

The di-methylated state of H3K_C27 has been shown to be linked to transcriptional repression and localizes to heterochromatin and enhancers (Barski et al., 2007; Wang et al., 2008). 20 factors were recruited whereas 23 factors were repelled from binding to H3K_C27me₂-modified chromatin (Figure 3.7 D, Table 3.1).

Several PHD finger proteins were recruited to H3K_C27me₂. Prominent examples are (i) the PHF21A protein, which is a component of the BHC co-repressor complex, (ii) the PHF15 (JADE2) protein, which is part of the HBO1 complex that displays histone H4 acetylation activity and (iii) PHF1, a protein, which is a component of the polycomb group. Remarkably, eight of the enriched factors are related to the spliceosome. SRRM1, MAGOH, EIF4A3, and SRRM2 are known to interact with the assembly intermediate C of the spliceosome complex, while the proteins SRSF6, PRPF40A, RNPS1 and SAP18 interact with the spliceosome at different stages of assembly. Interestingly, SAP18 is also part of the (chromatin) repressor complex SIN3, suggesting a crosstalk between the two processes. Known associations among the recruited proteins are shown in figure 3.7 E, upper panel.

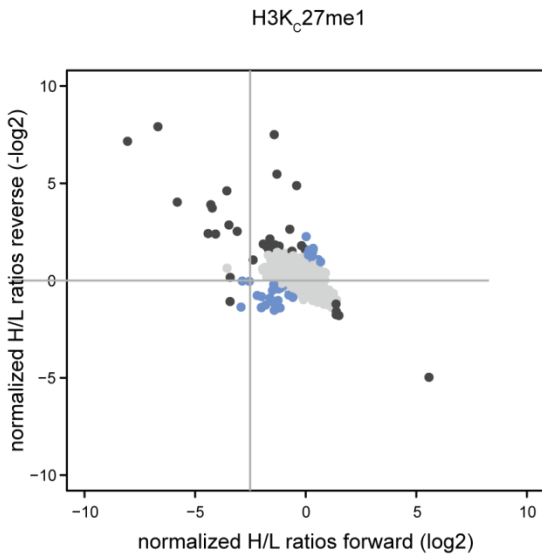
RNA splicing occurs either co-transcriptionally or immediately after transcription. Surprisingly, contrary to recruitment of spliceosomal factors, several subunits of the RNA polymerase I were excluded from binding to H3K_C27me₂. In addition, INO80, a component of the remodeling complex INO80, was also excluded from binding to H3K_C27me₂.

INO80 was additionally excluded from H3K9 mono- and tri- methylated chromatin. A similar binding profile was observed for the nuclear factors NFIB and E4F1 that were besides H3K9me₁/me₂/me₃ also excluded from H3K_C27me₂. These findings support a linkage of biological functions facilitated by H3K9me₁/me₂/me₃ and H3K_C27me₂ and indicate the exclusion of these factors from most heterochromatic regions. In general, the fact that several factors were regulated the same way by different modification states (Table 3.1) support the existence of a set of factors generally associated with the establishment and maintenance of certain chromatin stages.

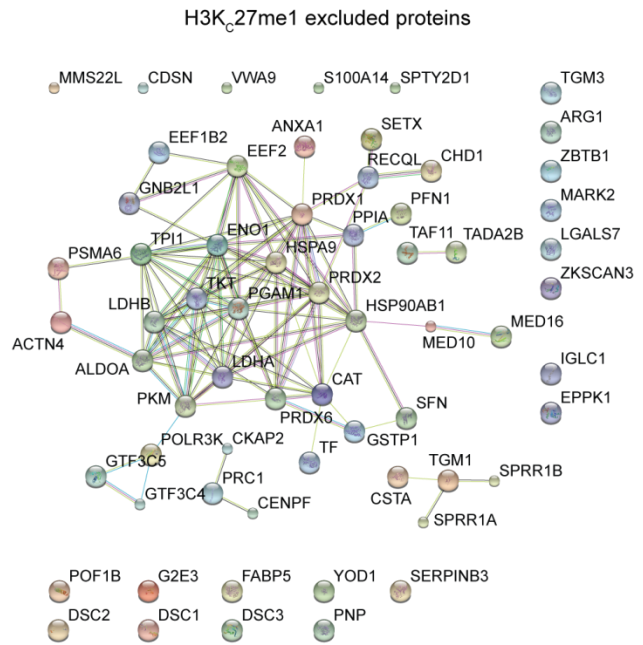
An example of proteins underlying opposed regulation by different modifications is CHAF1B, a member of the chromatin assembly complex 1 (CAF-1). The protein was excluded from binding to H3K_C27me₂ but enriched by H3K9me₃.

STRING analysis indicated that a protein complex comprised of the transcription factors SP1, SP3, ESRRA and NR2C1 was significantly depleted from H3K_C27me₂-modified chromatin arrays (Figure 3.7 E, lower panel). Another interaction among expelled proteins was identified between MLL3, a H3K4 methyltransferase, and the protein PAXIP1, which is involved in DNA damage response and transcriptional regulation (Wang et al., 2010).

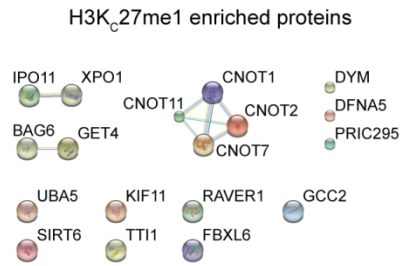
A)



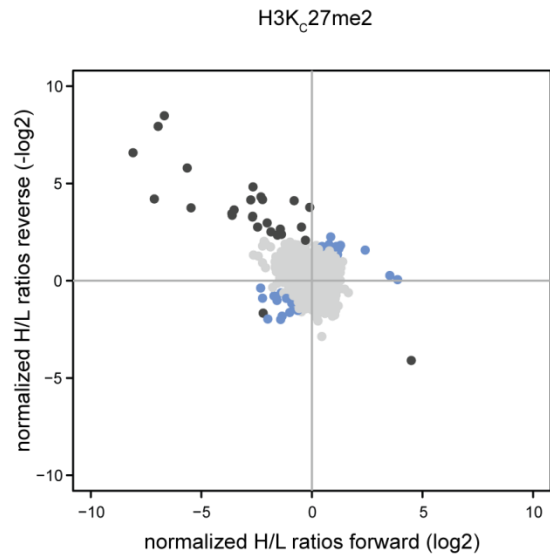
B)



C)



D)



E)

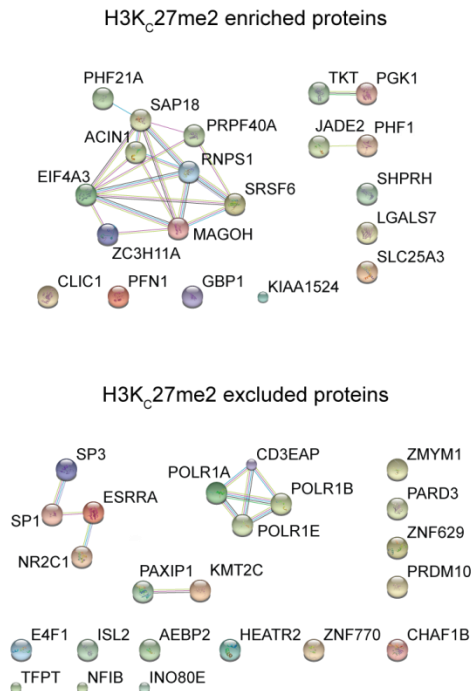


Figure 3.7 Protein-binding interactomes of H3K₂₇ mono- and di-methylated chromatin.

A) H3K₂₇me1 interactome. Proteins are plotted by their log₂ SILAC ratios of the forward experiment on the x-axis and the reverse experiment on the y-axis. Proteins that are significantly recruited or depleted from binding to chromatin by the modification are colored in blue. Proteins that were identified as false positive are marked in black. B) Protein-protein interaction network of proteins significantly excluded by H3K₂₇me1 predicted by STRING. C) STRING protein interactome of factors significantly recruited by H3K₂₇me1. D) H3K₂₇me2 interactome. Data are represented as described in A. E) STRING predicted protein interactome of factors significantly regulated by H3K₂₇me2. Recruited factors are shown in the upper part of the panel, whereas excluded factors are shown in the lower part of the panel.

H3K₂₇me3

The histone modification H3K27me3 is mainly present at silent promoters and is known to be related to gene silencing (Barski et al., 2007; Ringrose and Paro, 2004; Wang et al., 2008). Performing ChAP with H3K₂₇me3-modified chromatin arrays revealed that 22 proteins were recruited and 37 proteins were excluded from binding to this modification in the context of chromatin (Figure 3.8 A, Table 3.1).

Protein-protein interaction analysis using STRING predicted three multi subunit complexes among the proteins recruited to H3K₂₇me3. The complex encompassing the highest number of proteins is formed by several translation initiation factors, connected to the chromo domain protein CDYL2 (Figure 3.8 B, upper panel). A second complex is formed by the members of the ORC complex that was also shown to be recruited to H3K9me3-modified chromatin (Figure 3.8 B, upper panel). In a recent study, members of the ORC complex were shown to bind to both modification states, H3K9me3 and H3K27me3, in the context of mononucleosomes (Bartke et al., 2010). TOPBP1, which is required for DNA replication (Makiniemi et al., 2001), was also recruited but not shown to be in association with the ORC complex. A third complex was formed by the proteins PRC1, KIF2A and two members of the repressive polycomb group, CBX8 and PHF1 (Figure 3.8 B, upper panel). The polycomb group proteins are well studied and known to interact with H3K27 methylation states (van Kruijsbergen et al., 2015), therefore they were expected to bind to H3K27me3-modified chromatin.

The protein-binding interactome of H3K27me3 was already investigated in the context of mononucleosomes (Bartke et al., 2010). CBX8 was the only protein found to be recruited by H3K27me3-modified mononucleosomes and in the here presented study. No further overlap was observed, as in the context of mononucleosomes factors depleted from binding to H3K27me3 were not identified.

Factors like G2E3 and NOLC1 were identified to be excluded from binding to H3K₂₇me3-modified chromatin. These factors have been identified in more than one of the chromatin-binding interactomes, always in the context of exclusion from chromatin binding. Additionally, in contrast to H3K₂₇me2, factors related to the spliceosome are depleted by H3K₂₇me3, demonstrating an opposed protein-binding regulation by different modification states as

described for other factors before. Other expelled proteins are (i) transcriptional activators, (ii) factors functionally related to the cytoskeleton, (iii) proteins with kinase activity. The PHD finger proteins PHF8 and PH15 were found to be depleted as well.

The interactome of H3K₂₇me3 revealed the recruitment as well as repulsion of transcription factors from chromatin, which indicates the possibility of existing fine-tuning mechanisms of protein-binding properties mediated by a distinct PTM.

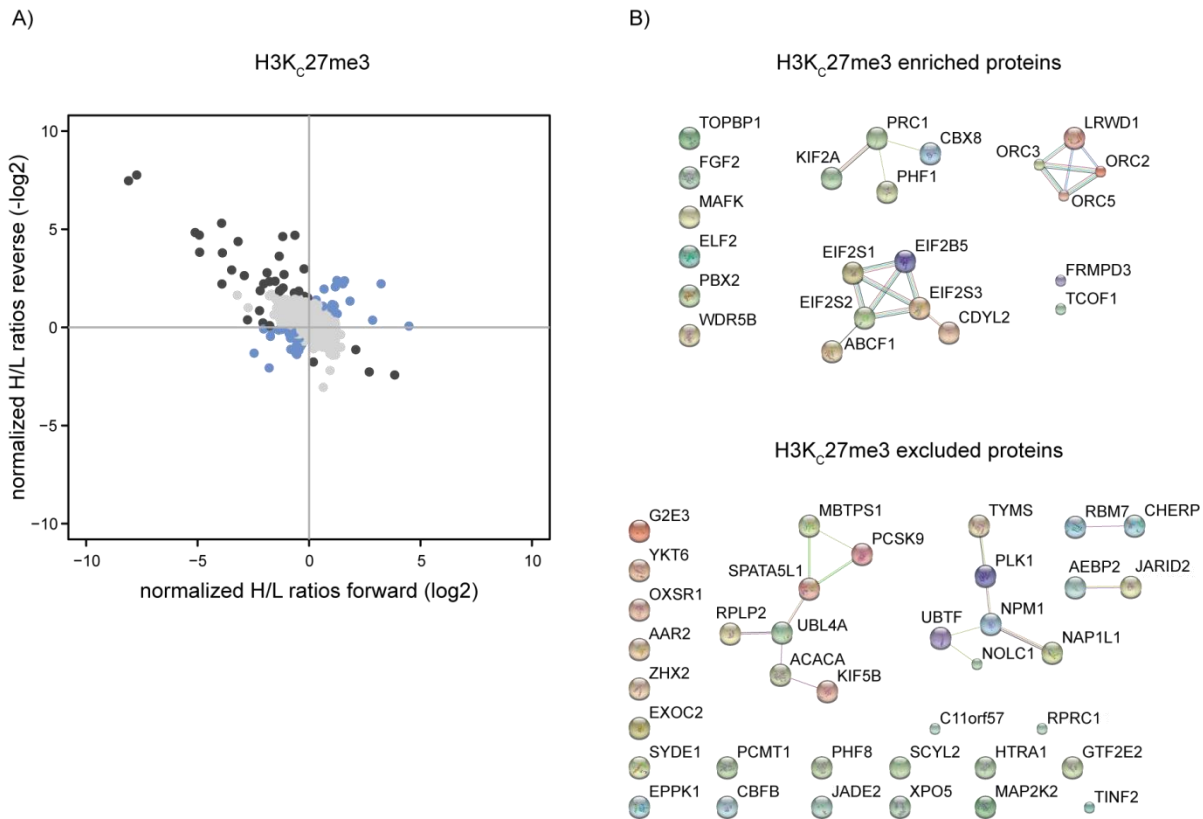


Figure 3.8 H3K₂₇me3 chromatin-binding interactome.

A) H3K₂₇me3 interactome. Data are represented as described in figure 3.7 A. B) STRING predicted protein-protein interaction networks of factors significantly recruited (upper part of the panel) and depleted (lower part of the panel) by H3K₂₇me3.

H4K20me1

Similar to all other investigated mono-methylation marks it was shown that H4K20me1 is linked to gene activation and co-localizes with H3K9me1 *in vivo* (Barski et al., 2007; Wang et al., 2008). 19 factors were enriched while 35 factors were excluded from binding to H4K20me1-modified chromatin (Figure 3.9 A, Table 3.1). Two components of the INO80 complex were recruited as well as the nucleosome assembly protein NAP1L4 and the RNA polymerase subunit gamma-1. Several factors involved in transcriptional repression were found to be recruited as well. Among them I identified PHF21A, HES1 and CIR1 (Figure 3.9 B, upper panel).

Protein-protein interaction analysis using STRING indicated that two complexes were depleted from binding to mono-methylated H4K20 (Figure 3.9 B, lower panel). The larger protein complex predicted by STRING consists of 18 proteins. 16 of these proteins are spliceosomal factors, mainly connected to the C complex. The second protein network includes members of the BRCA1-A complex, BARD1, BRCA1 and BRE, which are responsible for the maintenance of genome stability. The proteins ACACA and ATRIP have been shown to associate with this complex as well (Figure 3.9 B, lower panel). Altogether, the predicted STRING interactome connected 25 of the 35 excluded factors, suggesting most of the proteins to be linked to similar biological functionalities.

H4K20me3

Tri-methylated H4K20 has been shown to localize to heterochromatin and is connected to transcriptional repression (Barski et al., 2007; Schotta et al., 2004; Wang et al., 2008). Performing ChAP 32 factors were enriched and 11 factors excluded from binding to H4K20me3 chromatin (Figure 3.9 C, Table 3.1). The modification recruited the ORC complex together with its binding components LRWD1 and APEX1. With the exception of APEX1 the complex was also recruited to tri-methylated H3K9 and H3K_C27 (Table 3.1). The co-purification of the ORC complex with all three methylation sites has been shown before (Vermeulen et al., 2010). Additionally, STRING analysis predicted protein-protein associations between several recruited translation initiation factors (Figure 3.9 D, upper panel), which were also recruited to chromatin by H3K_C27me3. This complex showed additional association with two taxilin proteins connected to the hydrolase USP3, which is known to deubiquitinate monoubiquitinated target proteins, such as histone H2A and H2B. Several kinesin family members in association with DNAH8 and PHF1 formed a third complex predicted by STRING (Figure 3.9 D, upper panel).

H4K20me3 excluded several factors from binding to chromatin that have been shown to be excluded by other modification states investigated in this study as well. For instant, besides H4K20me3 all three methylation states of H3K9 excluded factors of the NFI gene family and the protein E4F1. The heterodimer BARD1-BRCA1 was excluded from binding to H4K20me1 as well as H4K20me3. Additionally, an identical regulation pattern was identified for a third protein, ANKRD12, which was regulated the same way by both introduced H4K20 methylation degrees.

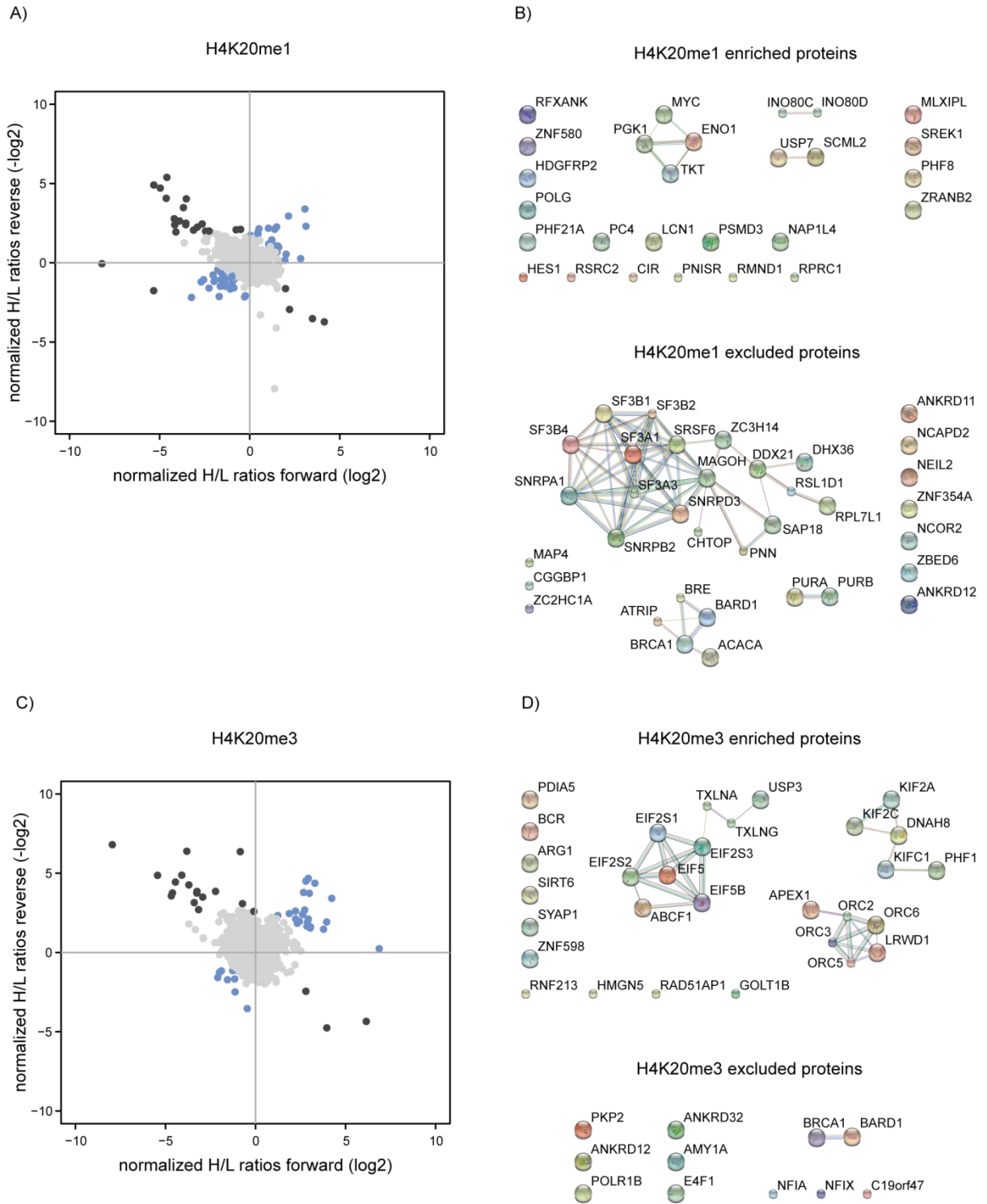


Figure 3.9 Chromatin-associated protein interactomes of histone H4K20 mono- and tri-methylation.

A) H4K20me1 interactome. Log₂ H/L ratios of proteins identified in the forward experiment are plotted on the x-axis and of the reverse experiment on the y-axis. Proteins that are significantly recruited or depleted from binding to chromatin arrays are colored in blue. False positive proteins are marked in black. B) STRING interactome of proteins significantly enriched (upper part of the panel) and excluded (lower part of the panel) by H4K20me1. C) H4K20me3 interactome. Data are represented as described in A. D) STRING protein-protein interaction network of proteins significantly recruited (upper part of the panel) and depleted (lower part of the panel) by H4K20me3.

H4R3me2

Arginine 3 of histone H4 is symmetrically di-methylated by PRMT5, a methyltransferase that is part of the repressive MB2/NURD complex. I found this complex recruited by H3K9me1 and depleted with H3K9me3 (Table 3.1). According to the literature, H4R3me2 was expected to be a mark associated with heterochromatin (Le Guezennec et al., 2006; Zhao et al., 2009). Nine factors were recruited to H4R3me2-modified chromatin and 11 factors were found to be excluded from binding (Figure 3.10 A, Table 3.1). Surprisingly, six of the recruited factors are DNA-directed polymerases, suggesting an association to transcriptional activity for H4R3me2. Additionally, two transcription factors and the protein SNC73 were recruited, supporting the assumption that this modification is rather associated with transcriptional activity and therefore unlikely to be linked to pericentric heterochromatin. This assumption is further supported by the fact that H4R3me2 excluded the proteins ACTL8, USP7, SCML2 and RAD51AP1 from binding to chromatin, as all proteins have been shown to be recruited by H3K9me3, a well-studied marker for pericentric heterochromatin.

H3Δ1-20

ChAP was performed with chromatin arrays containing truncated histones H3 that were missing the first 20 N-terminal amino acids (Figure 3.10 B). This experiment gave indications of protein binding to H3 influenced by the unmodified histone tail.

The experiment revealed 71 factors with a higher affinity to the wt H3, and thus these proteins were depleted from binding to chromatin when the N-terminus of H3 was missing (Table 3.1). USP7, DNMT1, ACTL8, and UHRF1 were found to be depleted by H3Δ1-20, which was not surprising as they have been demonstrated to be highly enriched in the context of several ChAP experiments with methylated H3 tails (Table 3.1). Other proteins excluded from binding to H3Δ1-20 were JADE3 (PHF16), several factors of the polycomb group, such as SUZ12, and the helicase SMARCA5. All these proteins are known to be associated with chromatin in the context of PTMs of the H3 N-terminus and therefore are not unlikely to be excluded from binding to H3Δ1-20 containing chromatin.

The proteins CDSN, IGKC and the histone methyltransferase KMT2C (MLL3) were found to be depleted by H3Δ1-20 as well as by chromatin arrays containing methylated H3. This finding indicates that these proteins preferentially associate with chromatin in presence of the unmodified H3 tail.

Interestingly, the N-terminal tail of H3 seems also to have a role in prevention of protein binding to chromatin. I identified 45 proteins that were significantly recruited to chromatin containing tailless histone H3 (Figure 3.10 B, Table 3.1). Most likely, these proteins showed enrichment in

consequence of a better accessibility to the globular domain of H3 or other parts of the nucleosome that might be sterically blocked by the presence of the H3 N-terminal tail.

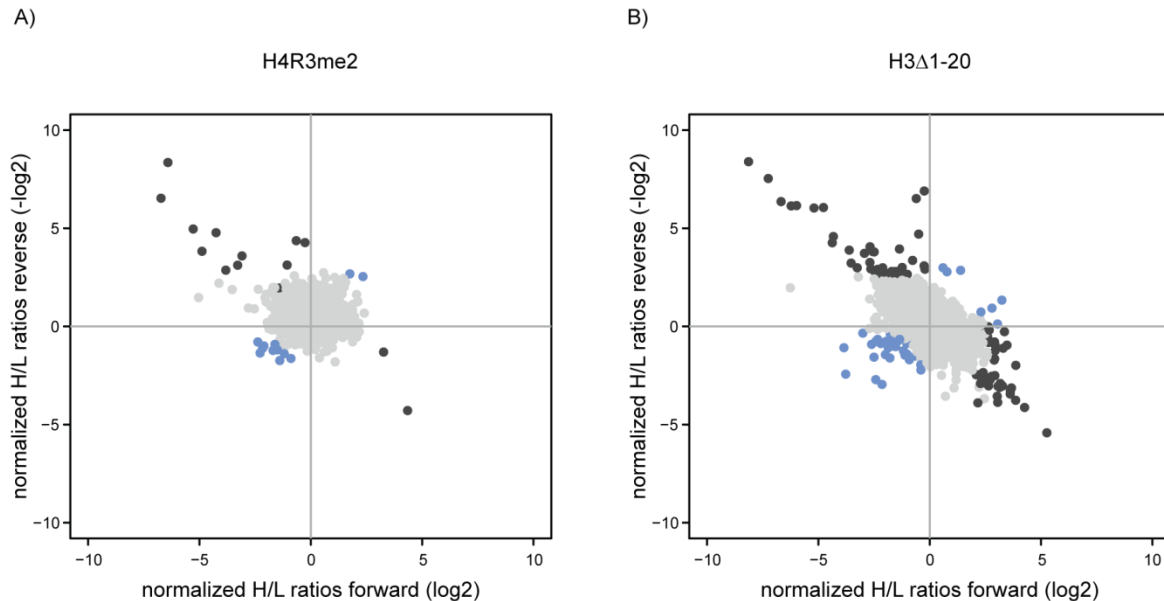


Figure 3.10 Chromatin-binding proteins regulated by H4R3me2 and H3Δ1-20.

Proteins are plotted by their \log_2 SILAC ratios of the forward experiment on the x-axis and the reverse experiment on the y-axis. Proteins that are significantly recruited or depleted from binding to chromatin by the modification are colored in blue. Proteins that were identified as false positive are marked in black. A) H4R3me2 protein-binding interactome. B) H3Δ1-20 interactome.

Methylated DNA (meCpG)

Chromatin arrays carrying methyl at position 5 of the cytosine pyrimidine ring within CpG stretches of the underlying DNA template recruited 42 factors and depleted 47 factors significantly from binding to nucleosomal arrays (Figure 3.11 A, Table 3.1). As DNA methylation has mainly been described as a repressive mark, I expected to find an overlap with factors also regulated by other repressive marks. Surprisingly, the only overlap of recruited factors was found between meCpG and H3K9me3. Among the overlapping nine factors were UHRF1, DNMT1, SCML2, USP7 and ACTL8. None of the factors regulated by H3K_c27me3 or H4K20me3 were regulated by methylated DNA (Table 3.1). 19 of the factors recruited by meCpG were predicted by the STRING database to be associated with each other (Figure 3.11 B). Among these proteins are chromo domain helicase binding proteins, metastasis associated 1 family members, the methyl CpG binding protein 2 and zinc finger proteins.

In the context of mononucleosomes affinity purification was performed with methylated DNA using a DNA template based on the 601 sequence (Bartke et al., 2010) similar to what was used in the here presented study. Five out of nine recruited factors that have been identified to be recruited in that study, namely, MECP2, MBD2, MTA2, CHD4, GATAD2A, were recruited by

DNA methylated oligonucleosomal arrays as well. The proteins MAX, USF2, BHLHB2 and USF1 were depleted from binding to both, mono- and oligonucleosomal arrays.

Protein-protein interaction analysis of excluded proteins indicated that only ten proteins were not associated with the predicted network (Figure 3.11 C). Several factors of the ING2 complex were excluded from binding including ING1 and its associated factor SFN. An additional complex was predicted by three proteins connected to the V-ATPase. DNA-dependent RNA polymerase and transcriptional activators were depleted from binding, supporting an association with gene silencing. The overlap observed among proteins excluded from meCpG and H3K9me3 binding is unique to polymerase subunits, namely POLR3F and POLR1C.

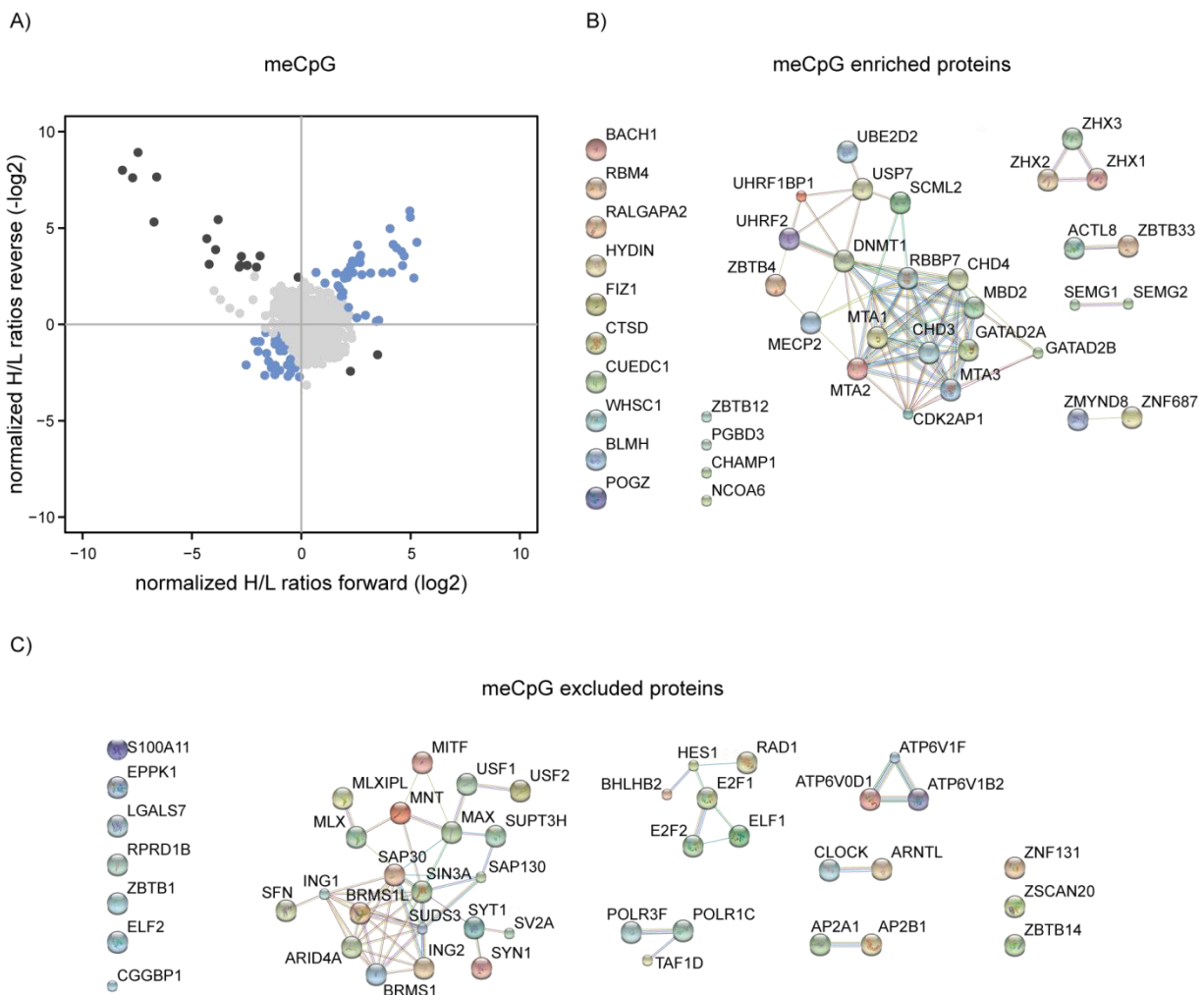


Figure 3.11 Protein-binding interactome of chromatin containing methylated DNA.

A) Protein-binding interactome of meCpG containing chromatin. Log₂ scaled H/L ratios of proteins identified in the forward experiment were plotted on the x-axis and the reverse experiment on the y-axis. Factors determined as significant outliers are marked in blue. False positives are marked in black. B) STRING protein-protein interaction analysis of factors significantly recruited by meCpG. C) Protein-protein network of proteins significantly depleted from meCpG chromatin predicted by STRING.

3.3.2 The methylation degrees of lysine residues affect protein binding to different extends

The functional significance of different methylation degrees of lysine residues within histones is still under discussion. The lower methylation degrees, mono- and di-methylation, might act as platforms that enable the establishment of the respective higher methylation state, leading to the tri-methylation mark that affects downstream processes. Alternatively, each of the three methylation degrees has their own biological significance and thus lead to distinct functional outcomes themselves.

In the following section I compared the protein-binding interactomes of different methylation degrees of H3K9, H3K27 and H4K20 to obtain more insights on how individual methylation degrees regulate chromatin-binding properties of individual factors and protein complexes.

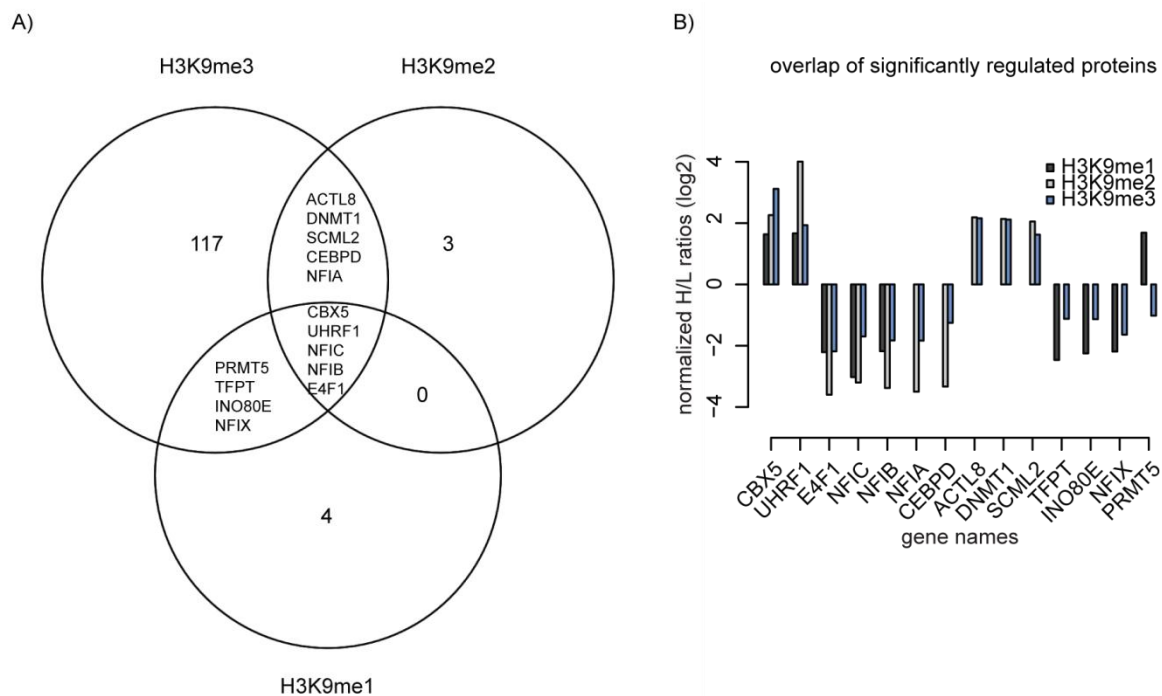


Figure 3.12 Overlap of proteins regulated by the methylation degrees of H3K9.

A) Venn diagram of all factors significantly regulated by mono-, di- and tri-methylation of H3K9. B) Bar plot of proteins regulated by at least two modification states of H3K9. Log₂ scaled SILAC ratios are plotted on the y-axis. Gene names are plotted on the x-axis.

Mono-, di- and tri-methylation of H3K9

Each of the methylation states of H3K9 recruited and repelled a specific set of factors from binding to the modified chromatin arrays (Figure 3.5, 3.6, Table 3.1). Remarkably, the number of factors regulated by H3K9me3 was 10 times higher than those regulated by H3K9me1 and

H3K9me2. In fact, mono- and di-methylation of H3K9 affected the binding behavior of 13 proteins, whereas H3K9me3 had an impact on 131 proteins. The overlap of significantly regulated factors is shown in figure 3.12 A. A set of five proteins, CBX5, UHRF1, NFIB, NFIC, and E4F1 was affected by all three modification states. These proteins were regulated the same way. They were either recruited to all modification states or excluded from binding to chromatin (Figure 3.12 B). Several proteins displayed an overlap between two of the modification states (Figure 3.12). With the exception of one factor, PRMT5, all proteins were regulated in the same direction by the overlapping methylation states. PRMT5 is the only factor found to display an opposed binding profile, as the protein was significantly recruited to H3K9me1 but repelled from binding to H3K9me3 (Figure 3.12 B).

Mono-, di- and tri-methylation of H3K27

The three different methylation states of H3K_C27 affected different protein interactomes, indicating the specificity of each modification state on protein-binding properties to chromatin (Figure 3.7, 3.8, Table 3.1). Notably, none of the modified lysine residues within histones investigated in this study demonstrated such a high divergence of protein binding regulated by different methylation degrees. The different H3K27 methylation states have been shown to localize to different genomic regions, with a greater diversity than H3K9 and H4K20 methylation states (Rosenfeld et al., 2009). Contrary to H3K9 methylation, not a single factor was found to be regulated by all three methylation states of H3K_C27 (Figure 3.13 A). An overlap of significantly regulated proteins was observed only between two of the H3K27 methylation states (Figure 3.13 A). The three factors, LGALS7, PFN1 and TKT were significantly affected by H3K_C27me1 and H3K_C27me2 and displayed an opposed binding profile to chromatin. All three were significantly repelled from binding to H3K_C27me1-modified chromatin while H3K_C27me2 significantly recruited these proteins (Table 3.1). JADE2 and the protein PRC1 also showed an opposed binding to chromatin in the context of different H3K27 methylation states. I also observed PHF1, AEBP2, EPPK1 and G2E3, to exhibit an opposed binding profile, thus, recruited and depleted from binding by two different H3K27 methylation states (Table 3.1).

Mono- and tri-methylation of H4K20

For H4K20 only mono- and tri-methylation states were investigated. Significantly regulated proteins of both modification degrees showed a high divergence (Figure 3.9, Table 3.1). An overlap was observed for only three factors, including ANKRD12, BRCA1 and BARD1 (Figure 3.13 B). All three proteins were excluded from binding independently of the H4K20 methylation state investigated.

Taken together, only a small set of common factors was observed to show the same binding profile to chromatin containing different modification degrees of a certain methylation site. The H3K9 methylation degrees showed an exceptionally high overlap of factors regulated the same way (Figure 3.12 A, Table 3.1). At least 70 % of H3K9me1 and of H3K9me2 regulated proteins were found to display the same binding profile as to H3K9me3. These findings suggest a functional link between the lower modification states and H3K9me3.

In contrast, diverse examples of proteins affected by the modification degrees of H3K27 and H4K20 highlighted the varying effect methylation degrees can have on the regulation of chromatin-binding interactomes that was not given for the H3K9 methylation states.

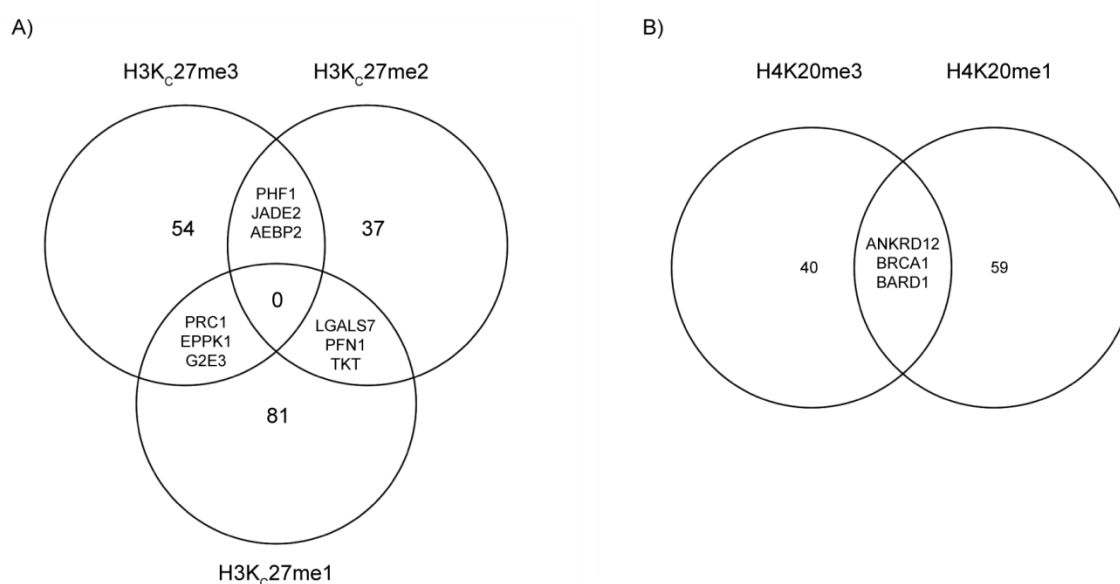


Figure 3.13 Overlap of significant proteins regulated by different methylation states.

A) Venn diagram of proteins significantly regulated by mono-, di- and tri-methylation of H3K_C27. B) Venn diagram of proteins significantly regulated by H4K20me1 and H4K20me3.

3.3.3 Two chromatin modifications *in trans* do not only constitute the sum of their single counterparts but demonstrate an independent impact on chromatin-binding interactomes

Crosstalk between histone modifications is referred to as the influence one modification has on the biological impact of a second one. It is mainly observed between two chemical modifications, resulting in changed chromatin-binding properties for a limited number of factors (Lee et al., 2010).

I hypothesized that crosstalk of two modifications does not only affect individual proteins but changes chromatin-binding properties of proteins on larger scale. Thus, I assumed that

chromatin modification crosstalk is a process of its own functional significance that extends the variety of chromatin regulating mechanisms.

In order to investigate chromatin modification crosstalk I extended the approach and applied ChAP by combining two posttranslational chromatin modifications. On one end I incorporated two modifications *in trans*, i.e. on different histones within one nucleosome. To this end we investigated the communication between modifications of H3 and H4. On the other end, I analyzed crosstalk between histone PTMs and DNA methylation in the context of chromatin arrays.

3.3.3.1 Chromatin arrays carrying a combination of H3K9me3 and H4K20me3 result in a specific protein interactome

To take advantage of the possibilities provided by the oligonucleosomal arrays I decided to investigate histone modification crosstalk between H3 and H4. Crosstalk *in trans* cannot be investigated using modified N-terminal peptides of histones, as it requires a nucleosomal context. Such analyses have not been done before on a proteomic scale. The combination of H3K9me3 and H4K20me3 was chosen as these modifications co-localize in pericentric heterochromatin *in vivo* and seem to be functionally linked (Mikkelsen et al., 2007; Schotta et al., 2004; Sims et al., 2006).

Performing ChAP using chromatin arrays carrying both modifications, H3K9me3 and H4K20me3 in combination (H3K9me3|H4K20me3), resulted in a protein interactome displaying 18 factors significantly enriched and 39 factors significantly repelled from binding to chromatin (Figure 3.14 A, Table 3.1). Interestingly, the interactome of the double modification was not just the sum or the average of the interactomes obtained with the individual modifications H3K9me3 and H4K20me3 (Figure 3.14 B). While all proteins specifically recruited by the double modification were also specifically recruited by at least one of the individual modifications, H3K9me3|H4K20me3 had a strong impact on exclusion of protein binding to chromatin. Here, overlapping proteins excluded from binding by H3K9me3|H4K20me3 and the individual modifications were rare and in fact only represented 5.1% of all excluded proteins. Thus, most factors excluded by the double modification were uniquely depleted from binding to chromatin only in presence of both, H3K9me3 and H4K20me3.

STRING analysis indicated that proteins repelled from binding to H3K9me3|H4K20me3 essentially clustered in four main protein complexes (Figure 3.14 C). Five members of the septin complex, four members of the HBO1 complex, responsible for the bulk of histone H4 acetylation, at least seven factors involved in splicing as well as several components of the BRCA1-A complex were identified. Interestingly, the components of the BRCA1-A complex, BARD1 and BRCA1, had been found to be excluded from binding to H4K20me3 as well (3.9 D, lower panel,

Table 3.1). These proteins are located in the background of the H3K9me3 interactome. In contrast, the members of the septin complex were significantly enriched by H3K9me3-modified chromatin (3.6 B, Table 3.1), demonstrating an opposed binding profile in the context of the double modification. These findings motivated me to investigate the binding profiles of proteins affected by all three modification states in more detail.

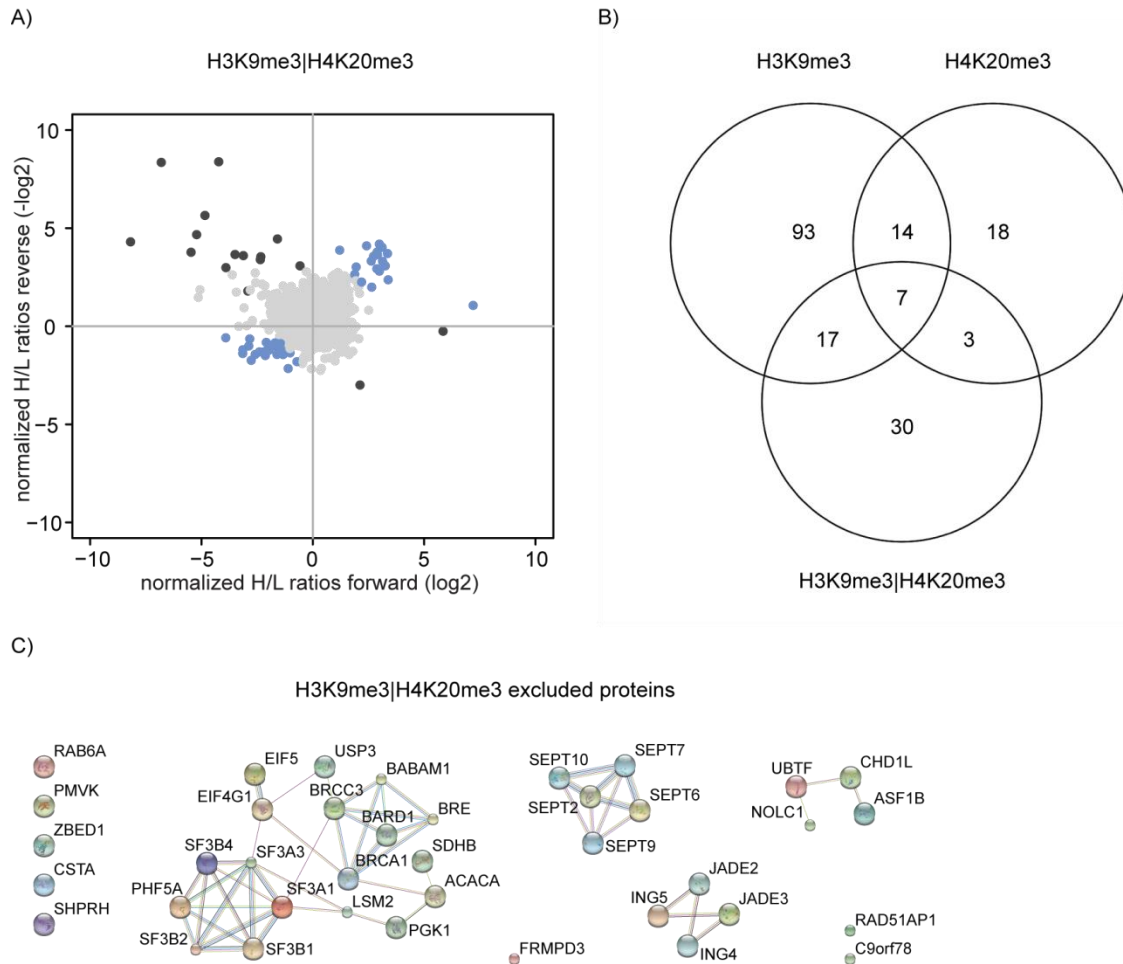


Figure 3.14 Protein-binding interactome of H3K9me3|H4K20me3-modified chromatin.

A) H3K9me3|H4K20me3 interactome. Proteins are plotted by their log₂ SILAC ratios of the forward experiment on the x-axis and the reverse experiment on the y-axis. Significantly recruited and depleted proteins are colored in blue. False positive proteins are marked in black. B) Venn diagram of factors significantly regulated by H3K9me3, H4K20me3 and H3K9me3|H4K20me3. C) STRING protein-protein interaction network of proteins significantly depleted from binding to H3K9me3|H4K20me3 chromatin.

3.3.3.2 The double modification H3K9me3|H4K20me3 indicates positive and negative crosstalk

The communication between different chemical chromatin modifications can occur on several levels. A modification can either promote or prevent the addition or the removal of a second modification. Furthermore, a modification can promote or prevent functional properties of a

second modification, e.g. the promotion or blockage of protein binding to chromatin and thereby functional consequences associated with these proteins.

The latter I investigated in the context of H3K9me3 and H4K20me3. Potential crosstalk was addressed by comparing the protein-binding profiles affected by the single modifications to the binding profile obtained with H3K9me3|H4K20me3. I started with proteins found to be significantly recruited or excluded from binding to the three different chromatin species (Table 3.1).

Considering only proteins significantly recruited by H3K9me3|H4K20me3 I found all proteins recruited by at least one of the single modifications as well (Figure 3.15 A). This finding suggested that protein recruitment in the context of H3K9me3 and H4K20me3 is defined by the presence of the individual modifications. The combination of both PTMs had no effect on protein recruitment independent H3K9me3 and H4K20me3.

Next, proteins significantly excluded from H3K9me3|H4K20me3 were compared to proteins excluded by the single modifications (Figure 3.15 B). Surprisingly, only 2 of the factors excluded by H3K9me3|H4K20me3 were also found to be excluded by one of the single modifications, in fact by H4K20me3.

When comparing H3K9me3|H4K20me3 excluded factors to significantly enriched factors of the two single modifications I found an overlap of 7 proteins (Figure 3.15 C). USP3, RAD51AP1 and EIF5 were recruited to both single modifications. Additionally, the members of the septin complex recruited by H3K9me3 were excluded by the double modification. These findings show for the first time that the presence of two modifications *in trans* result in an inverse binding pattern of proteins to chromatin. The combination of both PTMs had an opposing effect on protein binding in comparison to the individual modifications, H3K9me3 and H4K20me3.

I was wondering whether this effect was limited to a small number of H3K9me3|H4K20me3 excluded proteins. Therefore, I compared the SILAC ratios of all H3K9me3|H4K20me3 significantly excluded proteins with proteins identified in the experiments of H3K9me3- and H4K20me3-modified chromatin (Figure 3.16 A). Indeed, 32 of H3K9me3|H4K20me3 significantly excluded factors displayed an opposed binding affinity to chromatin compared to the individual modifications, indicating negative crosstalk between H3K9me3 and H4K20me3.

So far I only considered proteins that were significantly affected by the modifications. As I found proteins opposingly regulated by the double modification I was questioning whether this effect is limited to proteins excluded from binding. Proteins significantly recruited to H3K9me3|H4K20me3 showed no crosstalk. Therefore, I decided to extend the analysis and included proteins displaying an H/L ratio larger than $\log_2 1.5$ in the H3K9me3|H4K20me3 dataset. These proteins were not significantly recruited to the double modification but showed an

increased affinity to the modified chromatin arrays. 18 of these proteins displayed opposed binding compared to the single modifications (Figure 3.16 B), indicating also positive crosstalk between H3K9me3 and H4K20me3. Notably, almost all factors affected are in association with transcriptional activity, such as the DNA-directed RNA polymerase III subunits and the transcription factors NFIA and NFIC.

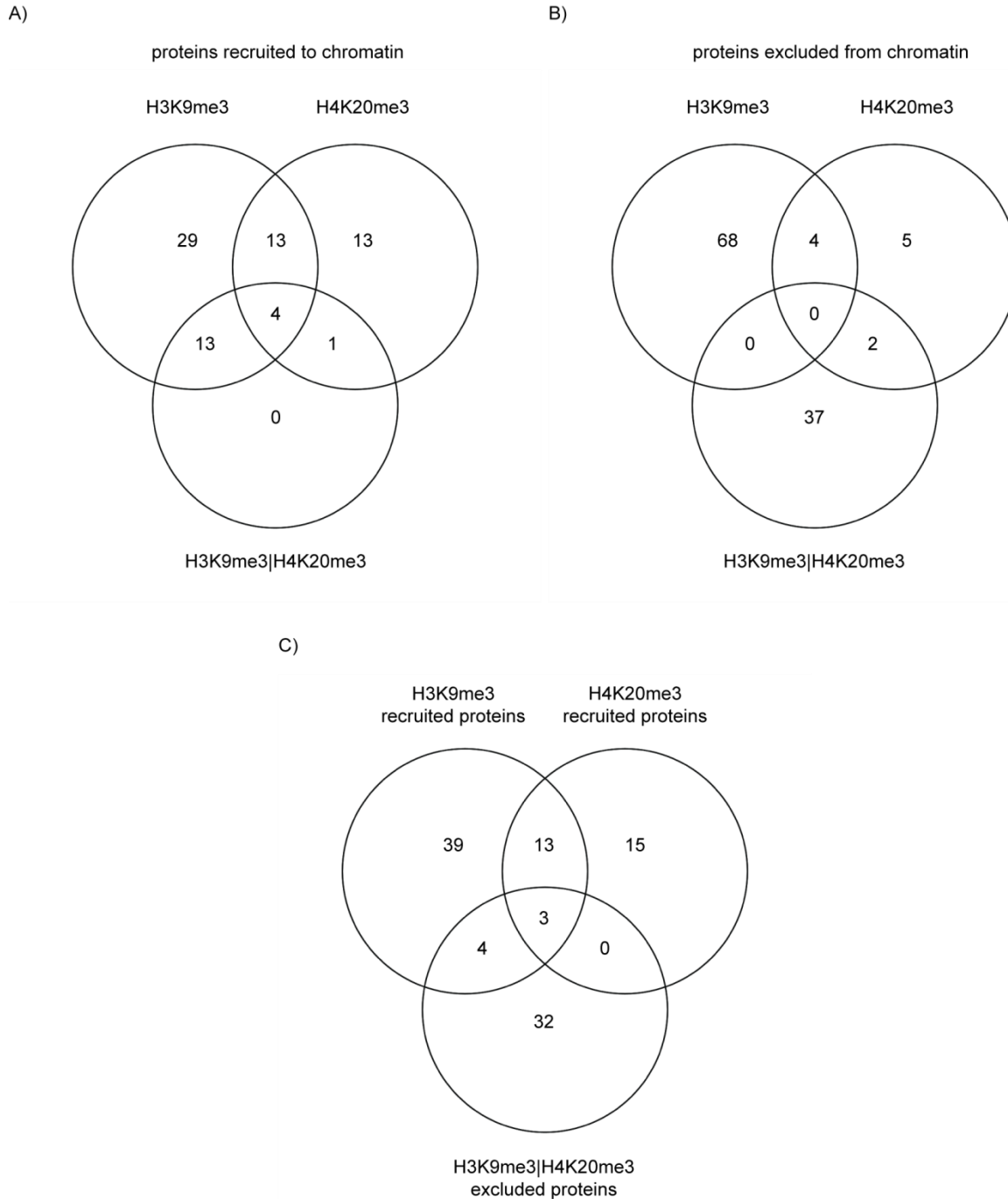


Figure 3.15 Comparison of significant protein binding to chromatin affected by H3K9me3, H4K20me3 and H3K9me3|H4K20me3.

A) Venn diagram of proteins recruited to chromatin by all three modification states. B) Venn diagram of proteins excluded from binding to chromatin by all three modification states. C) Venn diagram of proteins significantly recruited to H3K9me3 and H4K20me3 and excluded from binding to H3K9me3|H4K20me3 chromatin.

Taken together, the data showed three effects of H3K9me3 and H4K20me3 crosstalk. In the first category, the double modification had “no effect” on protein binding. “No effect” refers to the finding that the protein-binding properties were similar over all three modification patterns. In the second category, negative crosstalk between H3K9me3 and H4K20me3 was observed. Negative crosstalk was defined as decreased protein-binding affinities in the context of the double modification but enriched affinities with individual modifications. In the third category, H3K9me3 and H4K20me3 showed positive crosstalk. Positive crosstalk was given when the double modification displayed increased protein-binding affinities, while the individual modifications displayed decreased binding affinities.

3.3.3.3 Communication of histone and DNA modifications is indicated by positive and negative crosstalk

In many model organisms DNA methylation and histone PTMs have been found to be linked (Du et al., 2015) in the context of co-localization at certain chromatin regions, but also in terms of crosstalk. This made the combination of H3K9me3 and meCpG attractive to prove, whether the crosstalk observed between H3 and H4 is also given in the context of further combinations, i.e. DNA methylation and posttranslational histone modifications.

The analysis of the meCpG protein-binding interactome revealed an overlap of factors also significantly affected by the H3K9me3 interactome (3.17 B, Table 3.1). It appeared that a set of nine proteins is similarly regulated by H3K9me3 and meCpG, suggesting that these two modifications may be involved in the same regulation pathways. This argument was in agreement with a strong association between these chromatin modifications described in the literature (Du et al., 2015; Fuks et al., 2003; Lehnertz et al., 2003; Liu et al., 2013).

In consequence, I wanted to characterize in more detail, how the combination of both, histone posttranslational modifications and DNA methylation, regulates the chromatin-binding interactome. I wanted to investigate potential crosstalk, as seen with the combinatorial readout of H3 and H4 methylation. Therefore, ChAP experiments were performed using oligonucleosomal arrays carrying both H3K9me3 and CpG-methylated DNA.

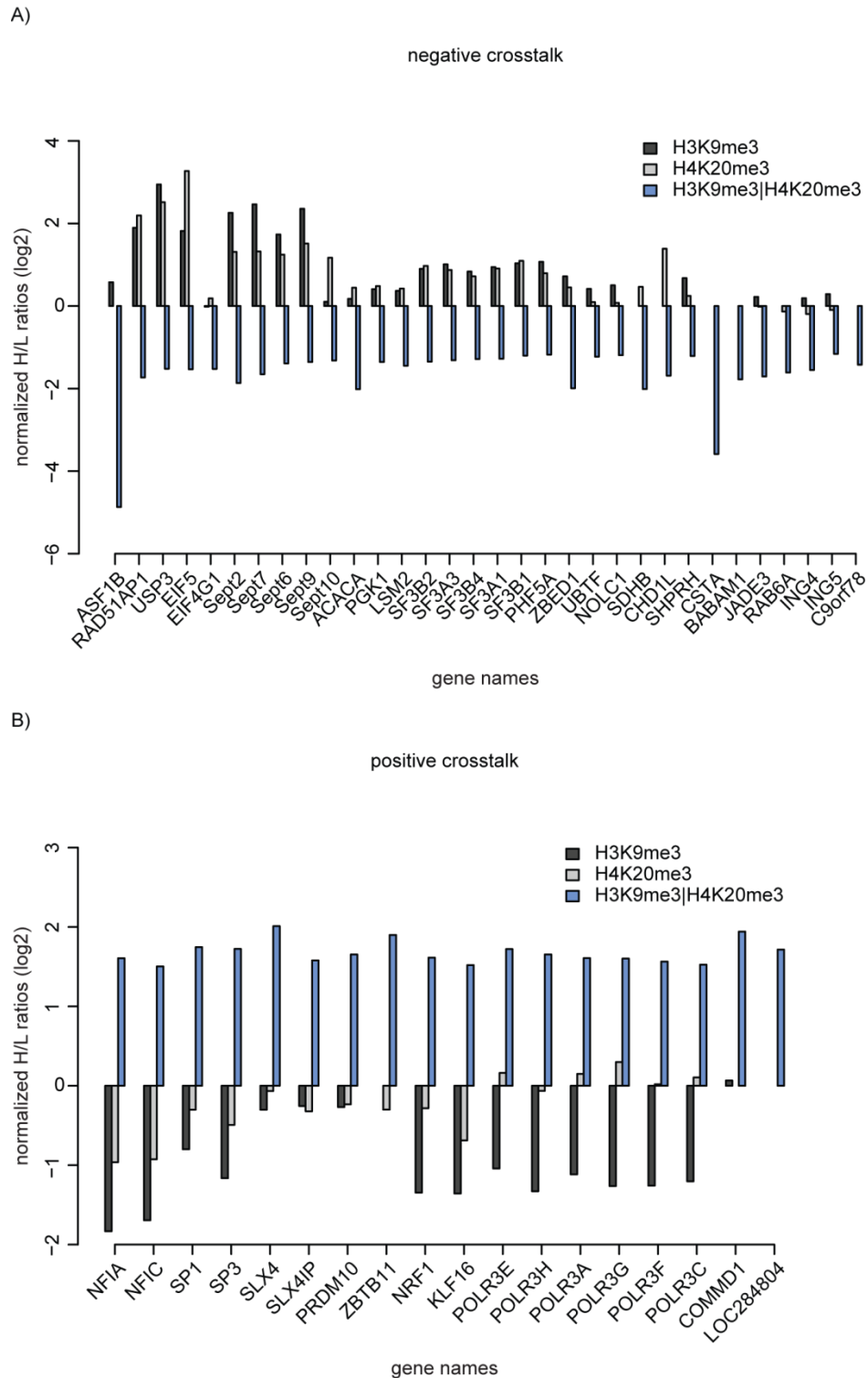


Figure 3.16 Crosstalk between H3K9me3 and H4K20me3.

A) Negative crosstalk between H3K9me3 and H4K20me3. Log₂ scaled SILAC ratios of proteins significantly repelled from binding and displaying negative crosstalk are indicated on the y-axis. There gene names are indicated on the x-axis. B) Positive crosstalk affected by H3K9me3 and H4K20me3. Proteins displaying an H/L ratio larger than log₂ 1.5 in the dataset of the double modification and displaying positive crosstalk are plotted as described in A).

The protein-binding interactome obtained from ChAP using chromatin arrays carrying both modifications, H3K9me3 and meCpG, in combination (H3K9me3|meCpG), displayed 7 factors significantly excluded from binding to chromatin but no factors significantly recruited (Figure 3.17 A, Table 3.1). The comparison of significantly affected proteins of all three datasets, H3K9me3, meCpG and the combination of both modifications, revealed an overlap of one factor, CDK2AP1, which was recruited to meCpG-modified chromatin and excluded from binding to H3K9me3|meCpG-modified chromatin (Figure 3.17 B). The opposed binding behavior to chromatin indicated negative crosstalk between meCpG and H3K9me3.

Despite the results of the significance test, an unusual high number of proteins displayed a higher affinity to H3K9me3|meCpG than to unmodified chromatin arrays, whereas the number of depleted proteins was similar to other ChAP-MS experiments (Figure 3.17 A). The possibility that this effect resulted from experimental variability is low, as only proteins displaying an increased H/L ratio are affected in both replicates. The high number of recruited proteins shifted the overall distribution towards high H/L ratios (Figure 3.17 A). With this large number of enriched proteins significant outliers could not be determined by the described significance test (section 3.2). However, considering the increased H/L ratios of both replicates, H3K9me3|meCpG-modified chromatin clearly displayed an impact on binding of a very large number of proteins to chromatin.

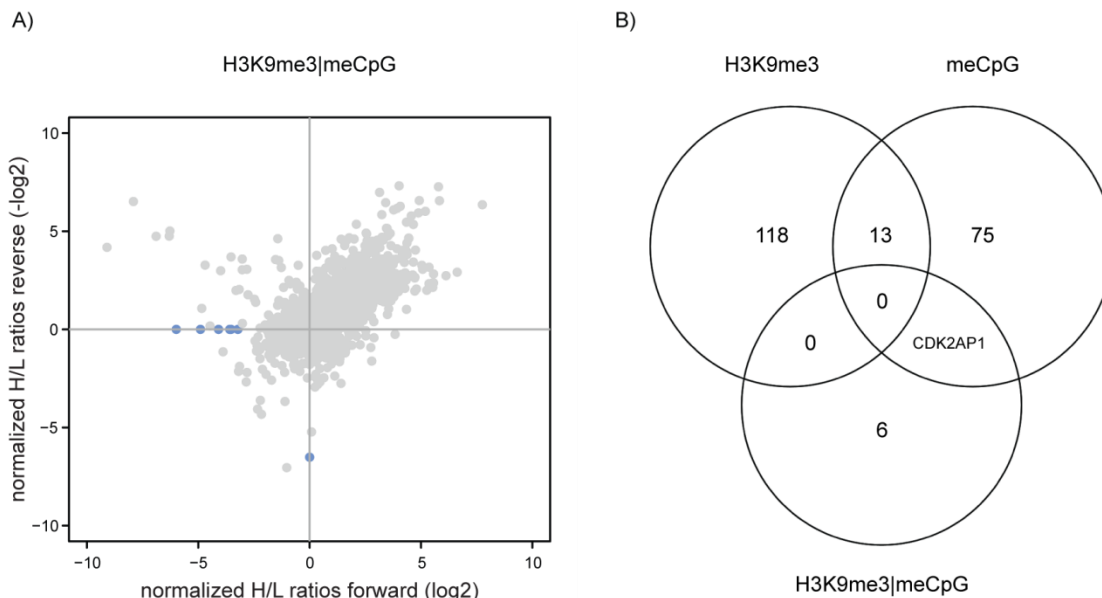


Figure 3.17 Chromatin-binding interactome of H3K9me3|meCpG in comparison to the individual modifications.

A) H3K9me3|meCpG interactome. Protein enrichment ratios are plotted as described in figure 3.14 A. B) Venn diagram of proteins significantly regulated by H3K9me3, CpG-methylated DNA and H3K9me3|meCpG-modified chromatin.

The combinatorial readout of H3K9me3|meCpG has already been performed in the context of mononucleosomes (Bartke et al., 2010). In comparison to the interactome revealed with chromatin arrays no overlap of significantly excluded factors was observed. Considering the 40 factors displaying the strongest enrichment to H3K9me3|meCpG chromatin and comparing these to the factors enriched with H3K9me3|meCpG mononucleosomes, an overlap of five factors was observed, namely UHRF1, CBX3, MBD2, MTA2 and CHD4. All five proteins have been shown to be enriched by at least one of the single modifications in the context of chromatin arrays as well.

In order to investigate positive and negative crosstalk I extended the analysis to proteins that showed no significance but increased and decreased binding affinities to H3K9me3|meCpG. Thus, all proteins displaying an H/L ratio larger than $\log_2 3.5$ and smaller than $\log_2 -2$ in the H3K9me3|meCpG dataset were included in the analysis.

Next to CDK2AP1, I found 14 more proteins showing a decreased binding affinity to H3K9me3|meCpG and an increased affinity to at least one of the individual modifications (Figure 3.18 A). The binding profiles of these proteins indicated negative crosstalk between H3K9me3 and meCpG. These proteins included USP3 and UBTF, which displayed negative crosstalk between both, H3K9me3 and meCpG and also H3K9me3 and H4K20me3. Other proteins affected were NUMA1, a component of the nuclear matrix, CENPF, which is involved in kinetochore function and chromosome segregation, the histone acetyltransferase KAT7 and the nucleolar protein NCL.

CpG-methylated DNA and H3K9me3 also showed positive crosstalk (Figure 3.18 B). The majority of affected factors are known to promote transcriptional activity. Interestingly, there was an overlap of 5 factors that displayed positive crosstalk between H3K9me3 and H4K20me3 as well. NFIA, NFIC, PRDM10 and the polymerase subunits POLR3A and POLR3H showed positive crosstalk between both investigated combinations of chromatin modifications. Additionally, three components of the chromatin remodeling complex INO80 and several transcription factors showed positive crosstalk in presence of H3K9me3 and methylated DNA.

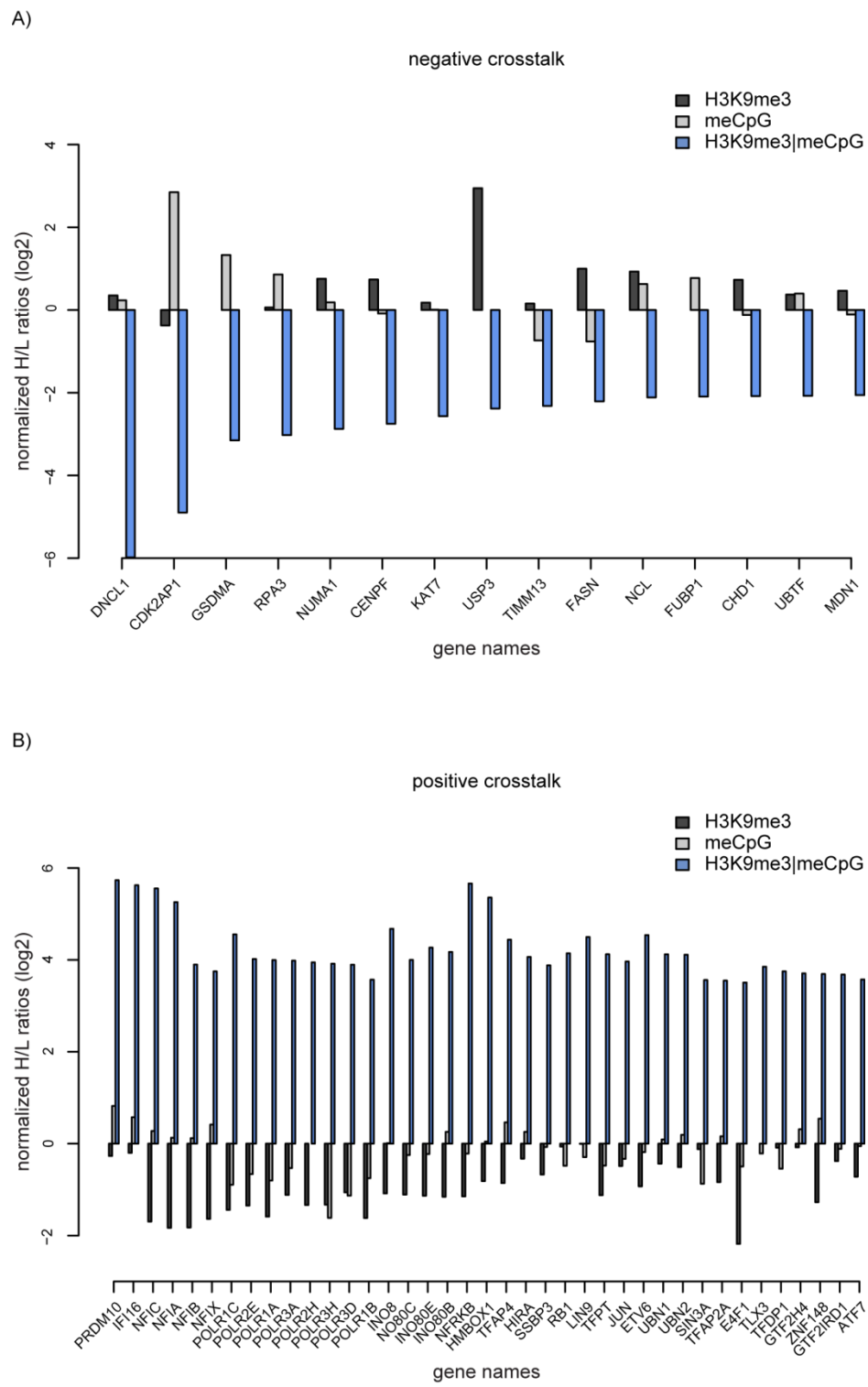


Figure 3.18 Crosstalk of H3K9me3 and CpG-methylated DNA.
A) Negative crosstalk of H3K9me3 and meCpG. Log₂ SILAC ratios are plotted on the y-axis. Gene names are plotted on the x-axis. Proteins with a SILAC ratio less than log₂ -2 were considered for the analysis. B) Positive crosstalk of H3K9me3 and meCpG. SILAC ratios and gene names were plotted as described for A). Proteins with a SILAC ratio greater than log₂ 3.5 were considered for the analysis.

3.4 Annotation analysis functionally correlates chromatin marks

High resolution maps based on genome wide analyses of chemical chromatin modifications provided insights into the distribution of these marks (Barski et al., 2007; Mikkelsen et al., 2007; Wang et al., 2008). Distinct modification patterns have been shown to correlate with certain chromatin states, leading to the assumption that a functional correlation between modifications exists. To investigate a functional correlation of chromatin modifications based on the protein interactomes (3.3), I applied gene annotation enrichment analyses. These analyses resulted in identification of overrepresented biological processes and functions related to individual chromatin modification marks.

As general strategy, significantly enriched and depleted proteins of a certain modification mark, introduced in chapter 3.3, were considered individually for gene annotation enrichment analyses. Gene names were mapped to associated biological annotations using the functional annotation chart of the online software package DAVID (Huang et al., 2009a, b). Results represented in the following chapter were obtained by using default statistic parameters of DAVID with the exception of the EASE score (enrichment) cutoff that was set to 0.05 to increase stringency. Annotation categories included in these analyses were GO terms for molecular functions (MF) and biological pathways. For the annotation of cellular pathways, the databases Panther_Pathways, KEGG_Pathway, BBID, Biocarta and Reactome_Pathway were screened. GO term annotation MF was performed including all terms. For each annotation term category two output lists per modification state were generated. One list of enriched annotation terms for recruited proteins (\uparrow) and one for proteins excluded from binding (\downarrow). Terms were only considered for enrichment analysis when at least two genes were assigned to a particular annotation term. Only significantly enriched annotation terms (matching the EASE score) were included in further analyses.

To enable global comparison of individual modification marks, the enriched annotation terms of recruited and depleted proteins of each chromatin modification mark were correlated by determining Pearson product-moment correlation coefficients (Pearson correlation coefficient). Modification marks showing a strong linear association of P-values of annotated terms displayed a high correlation coefficient. To compare annotated terms that were not enriched in each dataset missing P-values were set to 1.

Correlation coefficients obtained for enriched biological pathways are displayed in a heat map (Figure 3.19 A). Three clusters of strong correlation between modification states were observed. The largest cluster includes the datasets of recruited proteins of H3K9me3, H4K20me3, H3K₂₇me3 and H3K9me3|H4K20me3, indicating a comparable impact on biological functions.

All four chromatin modification marks have been shown to co-localize to pericentric heterochromatic regions and to be absent from transcriptionally active chromatin (Barski et al., 2007; Wang et al., 2008). Thus, functional similarities between these modifications might trigger related biological outcomes.

A second cluster of correlation coefficients was obtained for enriched biological pathways for the datasets of H3K9me3 and H3K₂₇me2 depleted and H4R3me2 recruited proteins (Figure 3.19 A). These findings were additionally supported by a strong correlation of H3K9me3 and H3K₂₇me2 depleted and H4R3me2 recruited proteins by annotation enrichment analyses of GO terms of MF (Figure 3.19 B). The results clearly indicated an antagonistic regulation of biological functions impacted by these modification marks. These biological functions are depleted by H3K9me3- and H3K₂₇me2-modified chromatin and given with H4R3me2-modified chromatin. On the individual protein level, I found several DNA-dependent RNA polymerase subunits excluded from binding to H3K9me3- and H3K₂₇me2-modified chromatin, while they were recruited by H4R3me2, suggesting transcriptional activity of chromatin regions marked with H4R3me2.

Another antagonistic regulation is highlighted by the third cluster of correlation coefficients obtained for enriched biological pathways of proteins recruited to chromatin modified at H3K₂₇me2 but depleted from binding to H3K9me3|H4K20me3 and H4K20me1 (Figure 3.19 A). All datasets were connected to the spliceosome, with the particularity that H3K₂₇me2 attracted, while the presence of H3K9me3|H4K20me3 and H4K20me1 excluded spliceosomal activity. This was attributed to the binding properties of several spliceosome associated factors that were shown to be opposingly regulated by H3K₂₇me2 compared to H3K9me3|H4K20me3 and H4K20me1 (section 3.3).

Correlation of P-values obtained from annotation term enrichment of GO terms MF pointed to a relation between the datasets of H4K20me1 recruited factors and factors depleted from binding to chromatin by H3K9me1/-me2, H3K₂₇me1/-me2 and H4K20me3 (Figure 3.19 B). These data suggested an antagonistic regulation of biological functions of H4K20me1 in comparison to the mono-methylation states of H3K9 and H3K27, which is surprising as they have been shown to be located at the same type of genomic regions (Barski et al., 2007; Mikkelsen et al., 2007; Rosenfeld et al., 2009; Wang et al., 2008). The correlation of opposed binding profiles of proteins to chromatin of H4K20me1 and the modifications H3K9me2, H3K₂₇me2 and H4K20me3 in contrast, is in agreement with previous findings. In genome wide distribution profiles H4K20me1 had its significance at transcriptional active sites whereas H3K9me2, H3K₂₇me2 and H4K20me3 accumulated at transcriptionally silenced regions (Barski et al., 2007; Mikkelsen et al., 2007; Rosenfeld et al., 2009; Wang et al., 2008).

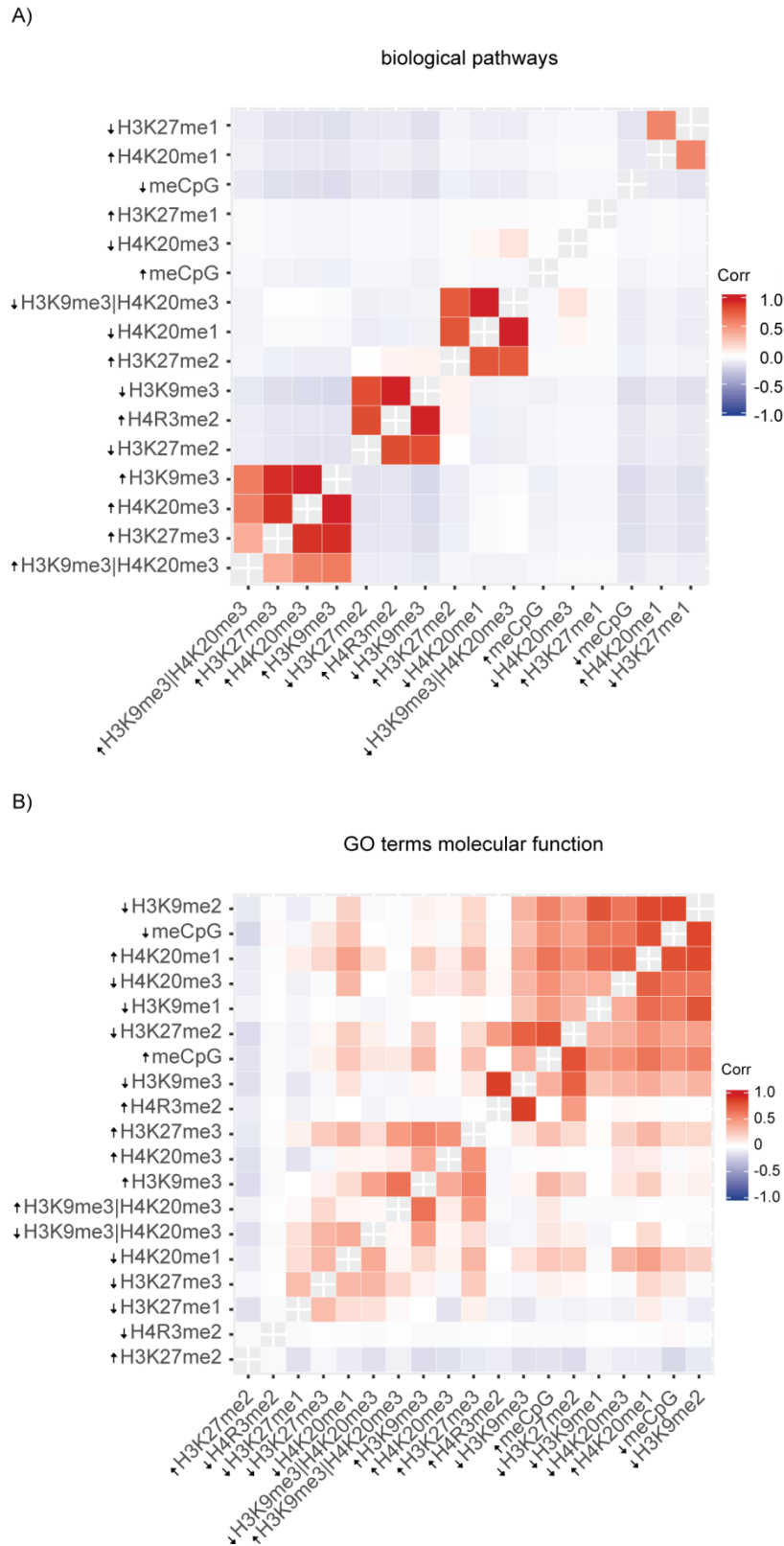


Figure 3.19 Annotation enrichment analyses of biological pathways and GO terms of molecular functions.

Significantly enriched (↑) and repelled proteins (↓) of all investigated chromatin modifications (section 3.3) were included in the analyses. Protein lists (enriched or excluded) resulted from the respective

modification states were applied to annotation term enrichment analysis. P-values of enriched annotation terms resulted from the modified Fisher's exact test performed by DAVID (Huang da et al., 2009a, b) were taken as input for calculation and visualization of Pearson correlation coefficients. P-values of annotated terms that displayed no significance were set to 1. A) Pearson correlation coefficients (Corr) of enriched pathway terms annotated to proteins recruited and excluded from binding to modified chromatin are listed on the y- and x-axis. Positive correlation is indicated in red shades while negative correlation is indicated in blue shades. B) Pearson correlation coefficients of enriched GO terms of molecular functions. Data are represented as described for A.

Notably, for both datasets of meCpG, proteins recruited and excluded from binding to chromatin, displayed a strong correlation (Figure 3.19 B). The functional classification of meCpG is still controversial. In recent years, more and more evidence has accumulated that connects methylated DNA with transcriptional activity (Hu et al., 2013; Jin et al., 2012; Wu et al., 2010). However, meCpG has been mainly correlated with transcriptional inactivity (Breiling and Lyko, 2015). In my analyses, proteins depleted as well as recruited to meCpG-modified chromatin were associated with similar biological functions, supporting a role for methylated DNA in the context of transcriptional activation as well as repression.

In summary, the results of gene annotation enrichment analyses of different modification patterns, based on significantly enriched and depleted proteins, suggested four groups of functionally related modification patterns. H3K9me3, H4K20me3, H3K_C27me3 and the double modification H3K9me3|H4K20me3 showed a strong correlation in the context of annotated pathway terms, suggesting a functional link between these modifications. Furthermore, the analyses indicated an opposing functional background for H4R3me2 compared to H3K9me3 and H3K_C27me2. The fact that H3K9me3 and H3K_C27me2 are associated with transcriptional inactivity and that H4R3me2 recruited several subunits of the polymerase (Table 3.1) strongly indicates that H4R3me2 is associated with transcriptional activity. The correlation analyses of enriched annotated terms also implied that H4K20me1 impacts protein binding to chromatin and associated functionalities opposed to H3K_C27me2. This finding is based on the inverse regulation of binding properties of spliceosomal factors (Table 3.1, Figure 3.7 E, upper panel, Figure 3.9 B, lower panel). Next to H3K_C27me2, also H3K9me1/-me2, H3K_C27me1 and H4K20me3 seem to trigger opposed biological functions than H4K20me1. A last group of modifications that seemed to be functionally linked, constituted H3K9me1/-me2, H4K20me3 and to a lesser extend meCpG. These modification states showed a correlation of terms annotated to proteins that were significantly depleted from binding to chromatin.

3.5 Protein-protein cross-linking of the chromatin-binding interactome

My results to that point demonstrated that chromatin affinity purification coupled to mass spectrometry is a powerful method for the identification of chromatin-associated factors. It allows determination of chromatin-binding interactomes in the context of different chromatin modifications and combinations thereof. The method allows studying the impact of chromatin modification patterns on recruitment and exclusion of protein binding to chromatin in addition to crosstalk effects of modifications (3.3). Yet, the approach does not provide information about factors that primarily bind to chromatin and factors that piggyback on other factors for association. To address the hierarchical binding pattern of chromatin-associated proteins we extended ChAP-MS by protein-protein cross-linking of the chromatin-binding interactome. A workflow combining ChAP, chemical cross-linking and mass spectrometry was established. This allowed the mapping of specific protein-protein interaction sites and thereby providing structural information of binding patterns of chromatin-associated factors and multi protein complexes.

3.5.1 Protein-protein cross-linking on assembled chromatin arrays

To identify physical interaction sites of specific chromatin-associated factors chemical cross-linking was applied. Chemical cross-linking results in permanent connections of protein interaction sites. Thus, proteins assembled in a non-covalent complex that are in close proximity, spanning the reach of the cross-linker, can be covalently linked.

In this study we used bis(sulfosuccinimidyl) suberate (BS3), a chemical cross-linker with a spacer arm length of 11.4 Å. It is a homobifunctional cross-linker that reacts with primary amines in the side chains of lysine (K) residues and the N-terminus of polypeptide chains resulting in stable amide bonds (Sinz, 2006). The appropriate BS3 concentration for efficient cross-linking was determined by titrating the cross-linker to affinity purified protein complexes bound to recombinant chromatin arrays (M. Nikolov, MPI-bpc). We used a concentration of 200 µM BS3 to sufficiently cross-link specific chromatin-binding interactomes.

The complete workflow established for cross-linking of chromatin-binding interactomes is summarized in figure 3.20. Briefly, specific chromatin-binding interactomes were obtained by ChAP. Biotin-tagged 12mer chromatin arrays were assembled as described before and validated for sufficient saturation levels (Figure 3.21). The ChAP workflow was extended by the addition of BS3 during affinity purification. Cross-linking was found to be most efficient when a small concentration of BS3 (20 µM) was already added to the pull down while chromatin arrays were still incubated with nuclear extract. The specific chromatin-binding interactomes were cross-

linked after protein-bound chromatin arrays were separated from nuclear extract and unspecific bound proteins were washed away.

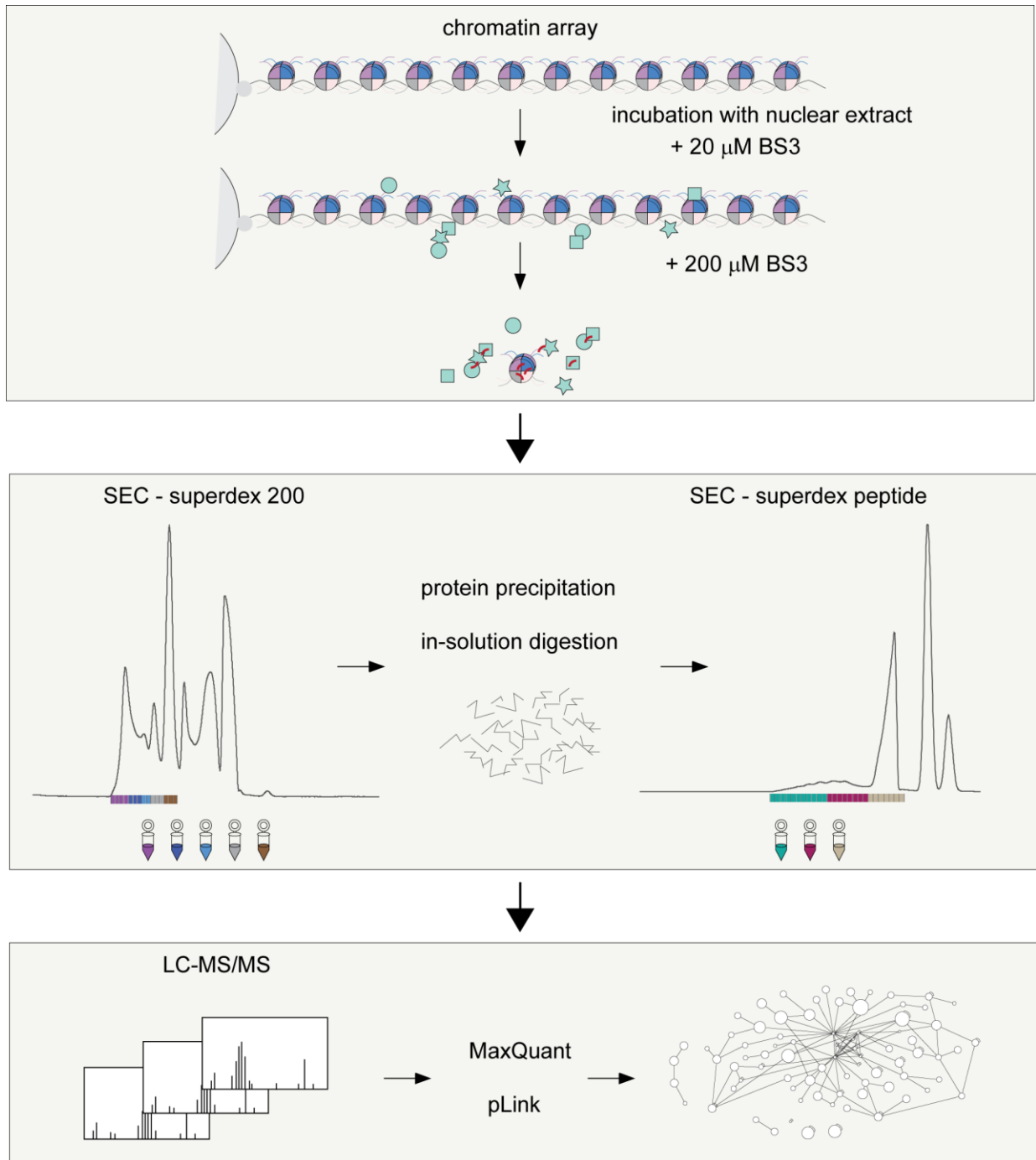


Figure 3.20 General workflow of protein-protein cross-link coupled with ChAP-MS.

An important step of the purification procedure is the removal of proteins cross-linked to streptavidin of the magnetic beads used for the chromatin pull-down. To exclude such cross-links and cross-links occurring between the streptavidin beads themselves from the analysis, the cross-linked chromatin-binding interactome was separated from magnetic beads. Therefore, the biotin-tagged DNA of the chromatin arrays was degraded by the nuclease benzonase resulting in the release of histone octamers and associated proteins from magnetic beads.

I prepared a negative control that allowed direct comparison of cross-linked with non-cross-linked samples, to prove that ChAP coupled with cross-linking enabled efficient and specific cross-linking on-beads of chromatin-associated proteins (Figure 3.22).

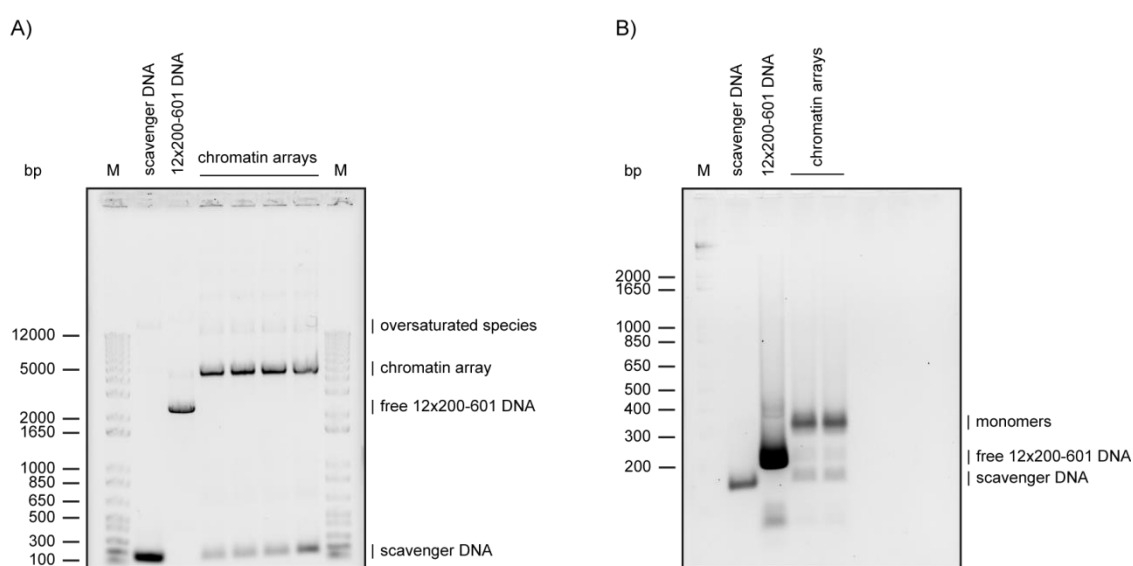


Figure 3.21 Quality control of reconstituted nucleosomal arrays.

(A) Agarose gel electrophoresis of free DNA and reconstituted nucleosomal 12mer arrays. (B) *Ava*I digestion of nucleosomal 12mer arrays shown in (A). The saturation level of the nucleosomal arrays is above 90%. M defines the DNA size marker.

The identification of cross-links from complex protein mixtures is challenging since cross-linked species are underrepresented. Therefore, it is necessary to reduce the complexity of the sample and enrich for cross-linked species before mass spectrometric measurements.

Separation of cross-linked and non-cross-linked species was performed by size exclusion chromatography (SEC) on the protein as well as on the peptide level. On the protein level, SEC elution profiles monitored by UV absorption at 215 nm featured 6 peaks (Figure 3.22 A, left panel). The protein mixture of chromatin associated factors was separated and several fractions were pooled according to eluted peaks. Cross-linked proteins eluted within the first three peaks from the column, visible by SDS-PAGE and Western blot analysis (Figure 3.22 A, middle and right panel). The analysis of elution fractions by SDS-PAGE validated the separation of proteins

according to their size and indicated the presence of large protein aggregates by a diffuse separation pattern of proteins (Figure 3.22 A, middle panel). The success of cross-linking was additionally demonstrated by upwards shifted signals of histone H3 as detected by Western blot analysis using an antibody against H3 (Figure 3.22 A, right panel). Fractions comprised of cross-linked material were used for further analyses. Proteins were digested in-solution using the protease trypsin.

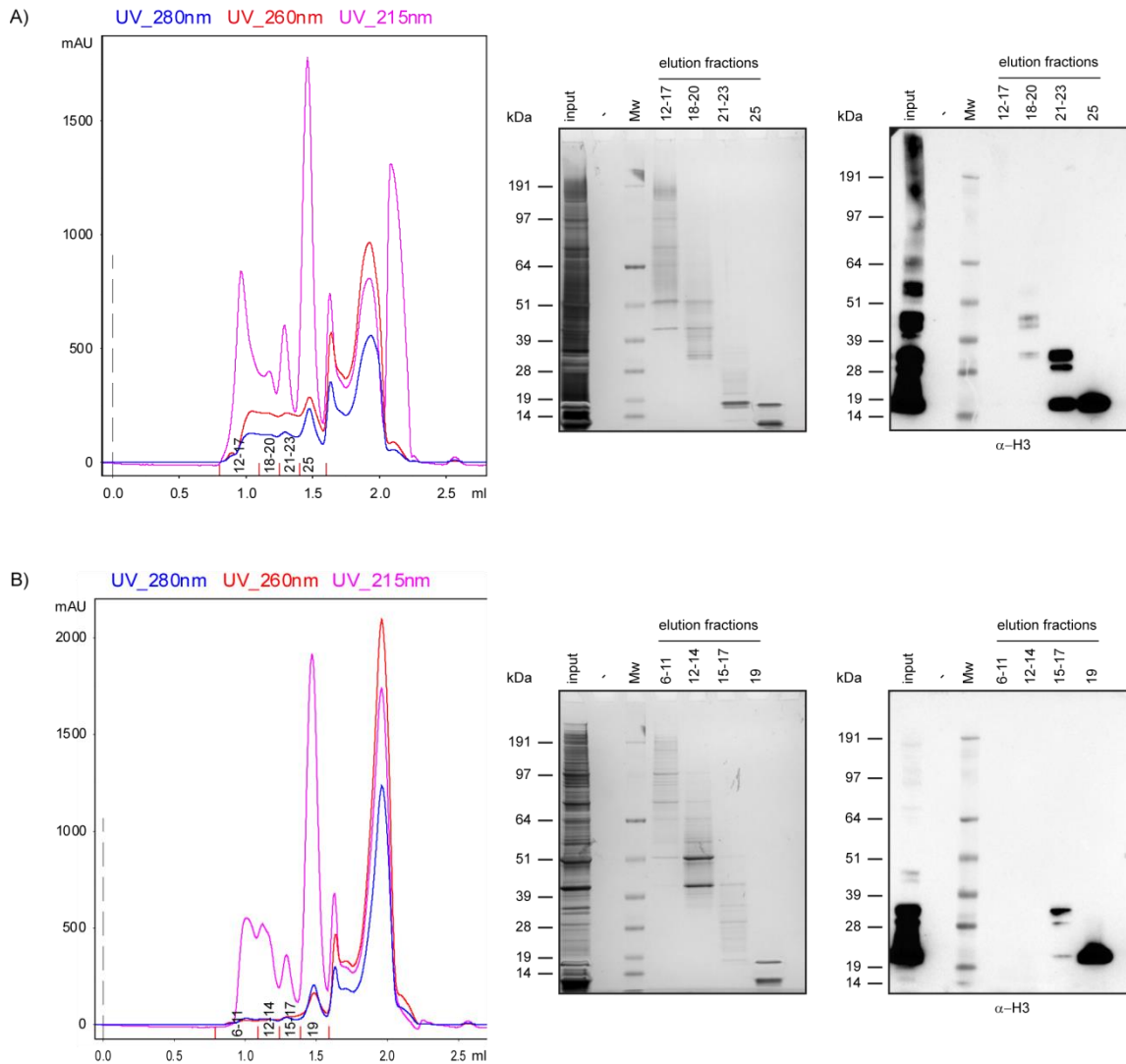


Figure 3.22 Enrichment for protein-protein cross-links by size exclusion chromatography (SEC).

Elution profiles (monitored by UV absorption) of proteins associated with unmodified chromatin that were separated according to their size using a Superdex 200 column (left panel). The protein contents of displayed peaks were analyzed by SDS-PAGE (middle panel) and Western blot analysis using apolyclonal α -H3 antibody (right panel). SDS-PAGE gels were stained with silver nitrate. Mw defines the molecular weight marker. (A) BS3 treated chromatin-associated proteins. (B) Negative control. All experimental steps of the negative control were performed equally to the cross-link experiment but the sample was not treated with BS3.

The negative control displayed a similar elution profile. Proteins started to elute with a shift of 0.1 ml (Figure 3.22 B, left panel), indicating larger protein complexes in the cross-linked samples. Elution fractions were pooled according to the cross-linked experiment (Figure 3.22 A, left panel).

The cross-linked protein samples treated with proteases were composed of three subpopulations of peptides, (i) cross-linked peptides, (ii) single peptide chains and (iii) non-cross-linked peptides. Single peptide chains can either be linked to a cross-linker reacted with water or ammonia (mono-links) or with a cross-linker that reacted with both ends of the same peptide (loop cross-links). Cross-linked peptides can occur as intramolecular bonds within one protein molecule (intra-cross-links) or as intermolecular bonds between two different protein molecules (inter-cross-links) (Figure 3.23 B) that are able to create high-molecular weight aggregates (Tran et al., 2016). Leitner and colleagues demonstrated that SEC at the peptide level of samples treated with cross-linkers facilitates the identification of cross-linked peptides by MS (Leitner et al., 2012). To ensure comprehensive detection of cross-linked peptides by MS, trypsinized protein mixtures were enriched for cross-linked peptides by another SEC step. The elution of peptides started with the void volume, indicating the presence of large peptides and the separation of cross-linked peptide aggregates from smaller peptide species. Elution fractions were pooled according to eluted peaks. Each of the resulting samples was analyzed by LC-MS/MS on a QExactive mass spectrometer. The majority of cross-linked peptides identified in the samples resulted from the first two peaks of the elution profile, additionally indicating a successful separation of cross-linked and non-cross-linked peptide species.

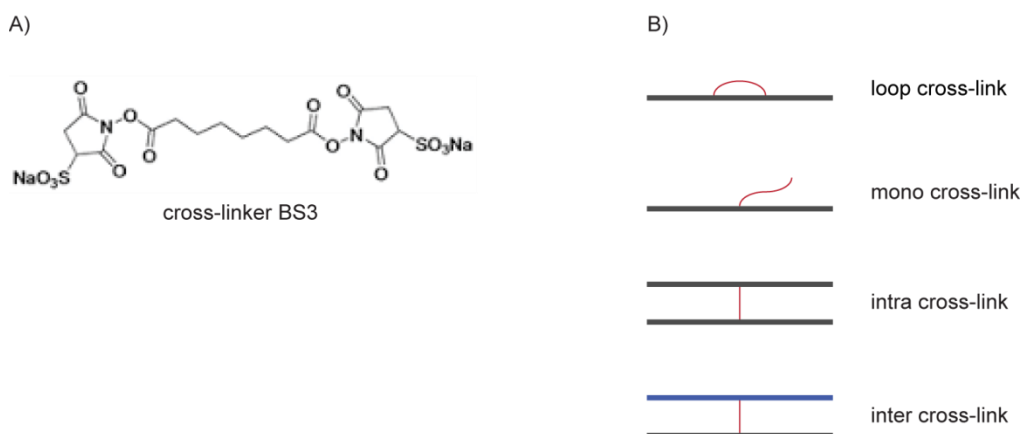


Figure 3.23 Chemical cross-links of peptides using the cross-linker BS3.

A) Chemical structure of bis(sulfosuccinimidyl) suberate (BS3). B) Cross-linked peptide species that can be obtained from proteolysis of cross-linked proteins (adapted from (Tran et al., 2016)).

3.5.2 Protein-protein cross-links of complex protein samples are specifically detected by the search algorithm pLink in combination with a reference database

One of the major challenges in XL-MS is the identification of cross-linked peptides and proteins. The main reason is due to the fact that obtained MS spectra do not correspond to one peptide but to all possible combinations of a cross-linked peptide pair. The consequence is a dramatic increase of the database used for the search associated with increased risk of random assignments and higher false discovery rates. These increases are exponential with the number of proteins included for the search. Thus, it is necessary to determine those proteins having a high probability to be detected in the search. These proteins will constitute the reference database.

To define a specific database, protein identification from MS measurements, were obtained using the MaxQuant software (Cox and Mann, 2008; Cox et al., 2011). Only proteins detected by at least 3 “razor and unique peptides” were included. The most abundant proteins were determined by dividing the number of detected peptide counts by their molecular weight.

To increase specificity of cross-link identification a FDR of 0.01 was set for the search using the pLink software (Purcell et al., 2007).

To test the specificity of protein-protein cross-link identification using pLink, I applied the algorithm to the mass spectra obtained from the negative control that should not show any cross-links. In total three pLink analyses were done, each time using a reference database including either the most 100 abundant proteins, the most 150 or all proteins identified in the sample. Depending on the size of the reference database that was used between 13 and 21 cross-links were identified. Three of those cross-links were also found to be positive identifications in the cross-linked sample and were excluded from further analysis. As the restriction settings for the detection of cross-linked peptides in combination with purification and enrichment strategies still included the identification of false positives the filtering stringency for the dataset was increased.

Therefore, the median of the pLink scores of all spectra that were identified for the negative control was determined. Spectra identified for the cross-linked samples displaying a score less than this median were excluded from further analyses. In total I found 0.6% false positive spectra that had to be excluded from further analysis. As the interest of investigation was focused on the identification of the binding hierarchy of proteins associated with chromatin, ambiguous spectra were generally excluded from analyses.

To further validate the specificity of cross-links I mapped the identified intra-cross-links of proteins against available crystal structures. The spacer arm of BS3 is 11.4-Å long. Lysines have a flexible side chain of 6-Å. Therefore, the two C α atoms of specifically cross-linked lysines

should be within a 24-Å distance. I selected the proteins Poly [ADP-ribose] polymerase 1 (PARP1) and DNA topoisomerase 2-alpha (TOP2A) from the dataset as they displayed several intra-cross-links and for both proteins crystal structures were available.

Using the Xlink Analyzer software (Kosinski et al., 2015; Pettersen et al., 2004) cross-links of PARP1 were mapped to the crystal structure of PARP1 with the PDB code 4DQY (Langelier et al., 2012). The crystal structure includes the amino acid sequences 6-91, 224-359, 531-575, 584-644, and 662-761. Hence, only 5 out of 20 cross-links could be mapped to the structure. All 5 cross links displayed a distance below 24-Å (Figure 3.24 A) validating the specificity of the detected cross-links. The same analysis was performed with the identified intra-cross-links of the human DNA topoisomerase 2-alpha (TOP2A), using the 2.9-Å resolution structure with the PDB code 4FM9 (Wendorff et al., 2012). This structure includes the amino acid sequences 433-1092 and 1124-1190 of the protein. Here, 5 out of 13 cross-links could be mapped. Only one cross-link featured a distance longer than 24-Å (33-Å). This cross-link is spanning the DNA-binding domain (Figure 3.24 B), which might be conformationally changed when TOP2A is not bound to DNA. Both, PARP1 and TOP2A are DNA-binding proteins (Figure 3.24) but have been identified to be cross-linked to proteins that themselves are directly cross-linked to one of the histones of the histone octamer (Figure 3.26 A).

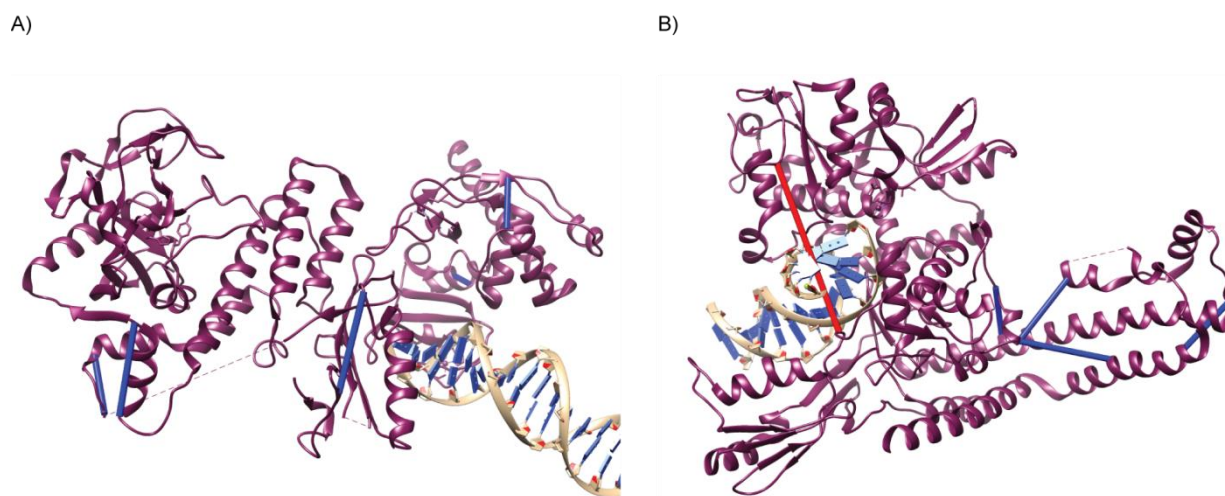


Figure 3.24 Protein-protein cross-links in the context of three-dimensional protein structures.

Identified intra-cross-links of Poly [ADP-ribose] polymerase 1 (PARP1) and DNA topoisomerase 2-alpha (TOP2A) were mapped to corresponding crystal structures using the Xlink Analyzer software (Kosinski et al., 2015; Pettersen et al., 2004). The tertiary protein structure is shown in purple. Cross-links are represented as blue bars when the distance is below 24-Å and in red when the distance is more than 24-Å. DNA bound by the protein is indicated by a golden alpha helix surrounding the sugar backbone in blue. A) Crystal structure of human PARP1 (PDB code 4DQY (Langelier et al., 2012)) demonstrating the position of all identified cross-links corresponding to the amino acid sequence of the shown structure. B) 2.9-Å resolution structure of human TOP2A (PDB code 4FM9 (Wendorff et al., 2012)). The position of identified cross-links corresponding to the amino acid sequence is shown.

In conclusion, with one exception I could demonstrate that all mapped cross-links lie within a 24-Å distance, suggesting that these cross-links are specific. The analysis of the negative control in parallel enabled the identification of false positive cross-links and therefore reduced the false positive rate. With the identification of only 0.6% false positive spectra, that were excluded from the dataset, and the mapping of identified cross-links to crystal structures as positive controls, it was shown that the stringency of data filtering is sufficient to provide reliable cross-link identifications.

3.6 ChAP-MS coupled with XL provides information of the binding hierarchy beyond primary binding proteins recruited to chromatin

ChAP-MS in combination with cross-linking was applied to unmodified nucleosomal arrays. For each experiment more than 200 protein species were identified. To ensure specificity and integrity of cross-link identification three different databases were used for pLink searches. The databases either comprised the top 100 abundant proteins, the top 150 or all positive identified proteins of one experiment. All together I obtained 1363 spectra with 521 different cross-link sites from two biological replicates. 45.7% of all identified cross-links localized to the nucleosome core particle.

Protein-protein cross-links of the core nucleosome

238 cross-links were identified within the core nucleosome, annotated in 792 spectra. These represented 464 inter- and 328 intra-cross-links. Most of these cross-links were mapped to the lysine-rich N-terminal or C-terminal tails of the histones (Figure 3.25 A). These terminal tails of the histones are thought to be unstructured, flexible and protrude outwards from the nucleosome core particle (Figure 1.1). It is known that the tails interact with the DNA wrapped around the nucleosome (Angelov et al., 2001; Mutskov et al., 1998) as well as with acidic patches of nucleosomes in close proximity (Davey et al., 2002; Dorigo et al., 2004; Luger et al., 1997). Hence, it is very likely that each of the histone tails can be temporarily in close proximity to each other.

As consequence, lysines far apart from each other in the crystal structure of the core nucleosome can come in close proximity and therefore can be cross-linked by BS3. This might explain that only 57 of all protein-protein cross-links identified within the core nucleosome were shorter than 24-Å whereas 174 of the cross-links were longer than 24-Å (Figure 3.25 B). In particular for histone H3 many cross-links were observed with a distance above 24-Å. As shown in figure 3.25 A cross-links to histone H3 were mainly found within the N-terminus of the protein. Similarly, for the histones H2A, H2B and H4 cross-link sites were concentrated within the

unstructured domains of the proteins (Figure 3.25 A). Therefore, no conclusion regarding structural information can be drawn from the cross-links obtained for the histone octamer.

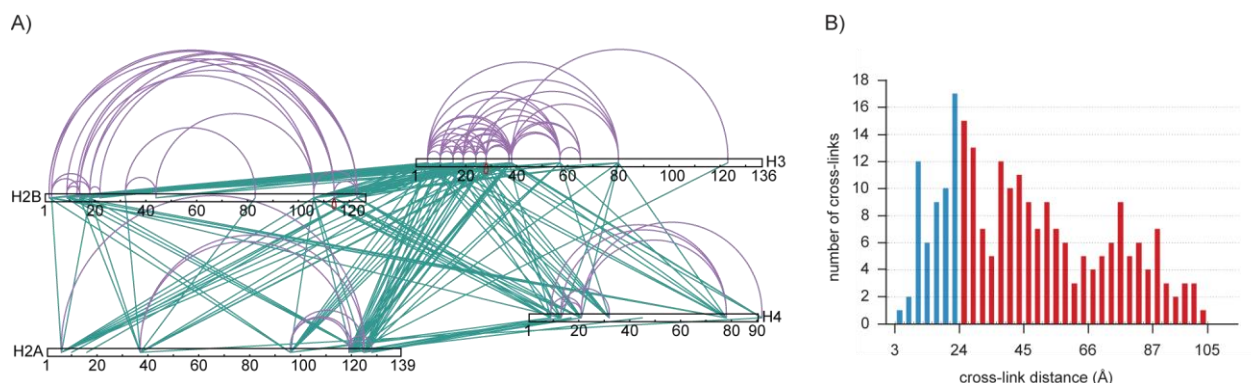


Figure 3.25 Protein-protein cross-link analysis of the core nucleosome.

A) Cross-links identified for the core nucleosome. The map was created using xiNet (Combe et al., 2015). Intra-cross-links are represented by purple lines. Inter-cross-links are colored in green. Red loops indicate that the same peptide was cross-linked to itself (multimerization). B) Histogram of cross-link distances of all identified cross-links within the four histone proteins mapped onto the crystal structure of the core nucleosome (PDB_1KX5 (Davey et al., 2002)) using the Xlink Analyzer software (Kosinski et al., 2015; Pettersen et al., 2004). Distances below 24-Å are labelled in blue. Distances longer than 24-Å are colored in red.

Protein-protein cross-links of chromatin-associated factors

In order to address the hierarchical binding patterning of chromatin-associated proteins and their physical interaction sites all filtered spectra of identified cross-links were uploaded to xiNET (Combe et al., 2015). The information of these spectra is summarized in a node-link diagram (Figure 3.26 A). 75 proteins were found to be cross-linked. Eight of them displayed only intra-cross-links, whereas all others exhibit at least one inter-cross-link (Figure 3.26 A). As cross-links covalently link proteins in close proximity, it was not surprising that 238 cross-links out of 521 different cross-link sites were identified within the histone octamer. 18 proteins showed a direct physical interaction to at least one of the core histones (Figure 3.26 A). Accordingly, these proteins were identified as primary binders to chromatin. A subset of these proteins was found to be cross-linked to a second protein. For instance, the protein HP1BP3 was cross-linked to histone H3 as well as to SMARCA5. Thus, this approach allowed the identification of secondary binders to the core nucleosome, in this case SMARCA5 (Figure 3.26 A). This result clearly demonstrate that SMARCA5 can be recruited to chromatin via HP1BP3 when not binding directly to chromatin by itself. The same is true for RPL6 and RREB1 as they were also identified to be cross-linked to HP1BP3. The protein SMARCA5 itself was cross-linked to a second protein, HSPA5, that can be considered as a tertiary binder to the core nucleosome. This protein again is connected to a multi subunit complex of proteins identified to be cross-linked to each other (Figure 3.26 A).

Not all cross-linked proteins could be assigned to one of the core histones. A protein complex of SNRNP200, HLTf, DNMT1, KDM2A, BAZ1B and HNRNPC was cross-linked to each other. In this case, the connector to chromatin was either not identified or the complex was directly bound to DNA via one of the proteins, for instance, such as DNMT1.

To prove that I cross-linked proteins with a biological background to chromatin that show expected functional associations, I investigated the 75 proteins identified to be cross-linked using the STRING database (Figure 3.26 B). Only six proteins were not included in the protein-protein map drawn by STRING, verifying the biological association of those proteins. The physical interaction sites of the four largest protein complexes of cross-linked proteins (Figure 3.26 A) were compared to the protein-protein interactions shown by STRING (Figure 3.26 B). Altogether, 17 protein-protein interactions overlapped between the two maps (Figure 3.26). For instance, the protein-protein interactions between the core histones as well as to histone H1 proteins have been also predicted by STRING. STRING also showed interactions between histone H2A and TOP2A. The complex of UBF1, TOP1, PARP1 and OGT1 was found in both maps as well.

Taken together, I could demonstrate that the established workflow is suitable for (i) identification of chromatin associated proteins, (ii) identification of physical interactions between proteins and proteins with the nucleosome, (iii) giving insights into the binding hierarchy of chromatin-associated proteins and their complexes, and (iv) identification of discrete interaction (cross-link) sites within proteins.

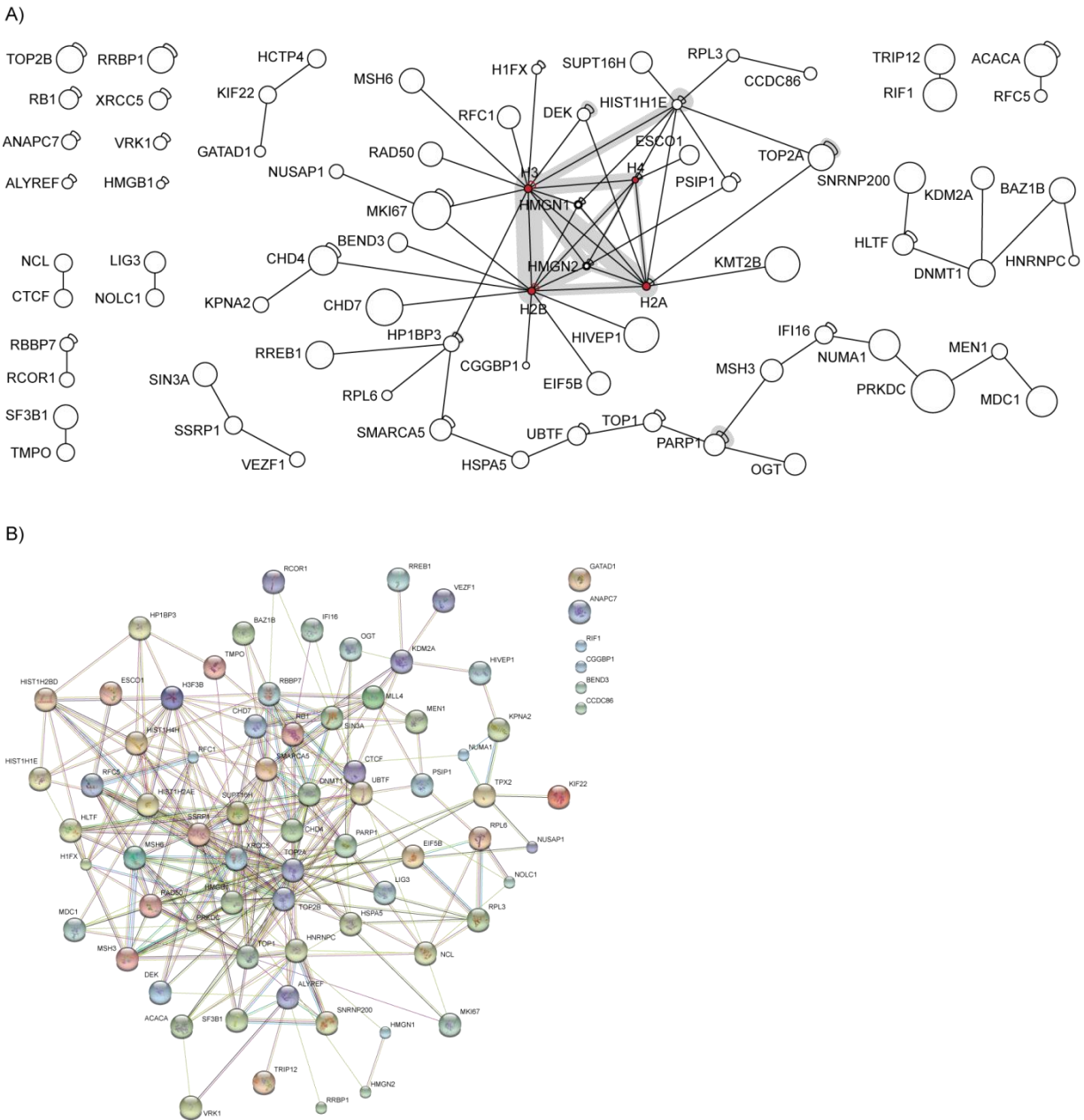


Figure 3.26 Interaction network of cross-linked factors associated with chromatin.

A) Node-link diagram of identified cross-links of chromatin-associated proteins. The network map was drawn in xiNET (Combe et al., 2015). Proteins are indicated by a node while cross-linked partner proteins are connected by a straight line between two nodes. Intra-cross-links are indicated by a loop connected to the respective node. Core histone proteins are marked in red. A large number of identified cross-links for a certain protein pair is roughly indicated by a grey shadow around the connection line between the two proteins. B) Protein-protein interaction network of known and predicted protein-protein interactions of the set of proteins found to be cross-linked in A. The network was generated by the database STRING (Jensen et al., 2009) using the default settings of the database.

4 Discussion

This study is based on a strategy combining recombinant homogenously modified chromatin arrays and quantitative mass spectrometry in order to characterize the protein-binding profiles in the context of 12 different histone modification patterns. This thesis work has established comprehensive lists of proteins whose binding to nucleosomes is regulated by histone modifications specifically found at heterochromatin. The raw proteomic data were analyzed by using a statistical approach, which allowed the comparison of different interactomes side by side. Conclusions were drawn regarding the biological functions, which are actively regulated by histone posttranslational modifications. Furthermore, the approach was extended by the incorporation of two different histone modifications in combination. For the first time, the crosstalk between one modification on H3 and one on H4 was analyzed on a global scale. Also, positive as well as negative crosstalk between histone modifications and methylated DNA was investigated. Lastly, a new approach combining chromatin affinity purification, cross-linking and mass spectrometry was developed to characterize the hierarchy of physical protein-protein interactions within a chromatin-binding interactome.

4.1 Individual chromatin-binding interactomes are comparable based on significant enrichment cutoffs

Affinity purification using 12mer chromatin arrays resulted in a high number of recruited proteins. In fact, I found between 1,117 and 2,785 proteins associated with chromatin depending on the incorporated modification. Most of the proteins were binding to the chromatin arrays independently of the histone modification state. To define these modification specific regulated factors the binding of proteins to modified and unmodified chromatin arrays were quantitatively determined and compared by their relative enrichment levels (SILAC ratio). Proteins significantly affected by a modification were identified by applying a statistical scoring function ("significance A") (Cox and Mann, 2008) to the distribution of log SILAC ratios in the corresponding experiment. Assuming a symmetric Gaussian distribution of variables (i.e. log SILAC ratios) this function calculates and scores the probability that a random individual variable has a defined distance from the mean of the distribution, which identifies it as a significant outlier. In average, approximately 3.5% of chromatin-bound proteins appeared to be directly affected by a chromatin modification.

The advantage of this statistical tool is a fold enrichment cutoff calculated based on the protein-binding properties of each individual experiment. Thus, significantly affected proteins could be determined independently of the variations introduced by the many experimental steps during sample preparation, stringently required for direct comparison of proteins affected by different histone modifications.

The drawback of this statistical tool is the requirement for a Gaussian distribution with small variance. In fact, when the binding properties of a very large number of proteins are affected by the underlying chromatin modification the set of SILAC ratios will spread out around the mean of the Gaussian distribution. In consequence, even the data indicate a very large number of proteins displaying an increased binding affinity to the modified chromatin arrays, no significant outlier will be determined.

Such a case was observed for H3K9me3|meCpG-modified chromatin. More than 30% of chromatin-bound proteins appeared to be affected by the double modification. In consequence, significant outliers could not be determined. Thus, for the interactome of the double modification H3K9m3|meCpG significantly recruited proteins could not be determined, (Figure 3.17 A). Hence, proteins recruited to H3K9m3|meCpG chromatin were not comparable to proteins significantly impacted by other modification states.

4.2 Interactomics of chromatin-binding proteins provide new insights into the functional impact of individual chromatin modifications

The complex language of posttranslational histone modifications is interpreted by proteins having reader domains. These histone-interacting domains, by recognizing their target sites, are thought to facilitate the recruitment of effector proteins and their connected biological functions to specific chromatin loci. The PTMs investigated here have a considerable impact on the composition of chromatin-bound proteins in the context of recombinant chromatin arrays. I investigated the protein-binding profiles in the context of 10 individual chromatin modifications. The first general conclusion from these datasets is the confirmation that PTMs as well as methylated DNA, promote and inhibit protein binding to chromatin. Second, independent of the methylation site and the degree of methylation, protein recruitment and exclusion from chromatin were regulated specifically, as shown by individual chromatin-binding interactomes for each modification. Although in the literature all investigated modification states were associated with heterochromatin, and therefore to be expected to have a certain overlap in regulated factors and functionality, a high diversity of factors was impacted by the different modification states. These factors display a limited set of common factors overlapping only partially between the different modification states. The observations support the notion of the existence of a histone PTM

language, guiding different biological functions, dependent on the underlying chromatin modification pattern, to specific genome loci (Jenuwein and Allis, 2001; Strahl and Allis, 2000).

Mono-, di- and tri-methylation of H3K9

The three methylation degrees of H3K9 show a strong contrast in numbers of proteins they significantly regulated. Considering all modification patterns investigated, H3K9me3 modulated the binding properties of the largest set of proteins. In contrast, H3K9me1/-me2 impacted the lowest number of proteins (Table 3.1). The finding indicates that lower methylation degrees of H3K9 affect chromatin in a more subtle way, likely regulating a more restricted set of functions.

At least 70 % of the proteins affected by H3K9me1/-me2 were also affected by H3K9me3 (Figure 3.12 A), suggesting a functional link between H3K9me1/-me2 and the tri-methylation state. The common functions regulated by these modifications might be heterochromatin establishment and maintenance. This assumption is supported by the facts that first, the methylation degrees of H3K9 are known to be involved in heterochromatin formation and maintenance (Loyola et al., 2009; Towbin et al., 2012). Second, in my study I found the well characterized human HP1 α (CBX5) and the heterochromatic protein UHRF1 to be recruited to chromatin by all three modification degrees (Table 3.1). Both proteins have been previously shown to bind to the different H3K9 methylation degrees and are primarily involved in heterochromatin formation and maintenance (Lachner et al., 2001; Nady et al., 2011).

Another common feature of the three H3K9 methylation degrees was the significant decrease in chromatin binding of transcription factors of the NFI gene family as well as E4F1 (Table 3.1). These factors correlated with the regulation pattern of polymerase subunits in the context of several modifications (Table 3.1). More precisely, the heterochromatic PTMs H3K9me3, H3K₂₇me2 and H4K20me3 repelled part of these transcription factors together with polymerase subunits from binding to chromatin (3.7 E, lower panel and 3.9 D, lower panel) whereas H4R3me2 specifically recruited polymerase subunits and NIFC (Table 3.1). Additionally, analysis using the database STRING predicted the factors of the NFI gene family to be in association with polymerase subunits (Figure 3.6 C). These findings suggest that the transcription factors of the NFI gene family work in the context of transcriptional activation and thereby support the heterochromatic context of the H3K9 methylation degrees.

The low number of H3K9me1/-me2 regulated proteins raises the question whether these modification degrees have their own biological significance or mainly function as substrates necessary for the establishment of H3K9me3 as suggested before (Loyola et al., 2009; Towbin et al., 2012). Indeed, my data implicate that H3K9me1/-me2 do not autonomously promote functional outcomes and are related to H3K9me3. This supports the suggestion of H3K9me1/-

me2 being substrates for establishment of H3K9 tri-methylation. However, the independent genomic distribution of the modification degrees indicates the contribution to various chromatin related functions (Barski et al., 2007; Mikkelsen et al., 2007). Therefore, it cannot be ruled out that the lower methylation states of H3K9 act in concert with other modification sites to promote downstream functions at chromatin regions free of H3K9me3.

Mono-, di- and tri-methylation of H3K_C27

In comparison to H3K9, the characteristics of the three methylation degrees of H3K_C27 feature three main distinctions. First, there was not a single factor found to be commonly regulated by all three methylation degrees of H3K_C27 (Figure 3.13 A). Second, the factors regulated by each of the modification degrees were highly divergent in functionality. Third, chromatin binding of several factors was regulated in an antagonistic manner by two of the H3K_C27 methylation degrees (Table 3.1). These observations indicate a specific biological significance for each of the methylation degrees. This point of view agrees with previous findings demonstrating that the different H3K27 methylation states are located in different regions of the genome, with a diversity much greater than for H3K9 and H4K20 methylation marks (Rosenfeld et al., 2009).

The mono-methylation of H3K_C27 recruited several components of the CCR4-NOT complex, which has been shown to be involved in transcription initiation and elongation (Collart, 2016). In this context the complex was shown to control H3 and H4 acetylation and H3K4 tri-methylation (Peng et al., 2008), indicating a role for H3K27me1 in transcriptional activation. This would be in agreement with the fact that H3K27me1 is enriched in gene bodies of actively transcribed genes (Barski et al., 2007; Wang et al., 2008). The CCR4-NOT complex has also been shown to act in the cytoplasm, regulating mRNA translation as well as degradation. Surprisingly, there are three more proteins recruited to H3K_C27me1 that also show cytoplasmic localization. These proteins seem to be associated with a correct organization of the Golgi apparatus and transport processes between the Golgi apparatus and the endoplasmic reticulum or endosomes (Dimitrov et al., 2009; Mariappan et al., 2010; Reddy et al., 2006). It is unclear how these proteins are related to H3K27me1. These proteins might represent cytosolic contaminants found in the nuclear extracts due to the method used for the preparation of the extracts (Dignam et al., 1983). I think this is not very likely. First of all, light and heavy nuclear extracts were prepared individually. Secondly, for each of the two extracts three individual preparations were performed and pooled in order to average potential abnormalities introduced by experimental handling. Thirdly, some of the cytoplasmic proteins identified in this study were also identified by affinity purification in the context of N-terminal histone peptides using HeLa nuclear extracts (Vermeulen et al., 2010). From my point of view these facts demonstrate the improbability of contaminants in

the nuclear extracts used for this study and support the discovery of so far cytoplasmic classified proteins as positive identifications.

Remarkably, 40% of the proteins recruited to H3K₂₇me₂ chromatin are spliceosomal proteins. A link between regulation of alternative splicing and chromatin modifications has already been described (Luco et al., 2011; Luco et al., 2010; Zhou et al., 2014). The nucleosome density and consequently the presence of histone modifications was found to be higher over exons compared to introns (Schwartz et al., 2009; Tilgner et al., 2009). The presence of histone modifications is thought to locally impact alternative splicing by either, recruitment of splicing factors to the transcription site, or by marking alternative exons and/or the surrounding chromosomal regions and influencing the elongation rate of polymerase II and therefore splice site choices. The presence of H3K9me₃ and HP1 γ was shown to slow down polymerase II resulting in increased inclusion of transcribed alternative exons (Saint-Andre et al., 2011). In contrast, hyperacetylation of H3 and H4 increased the elongation rate of polymerase II and favored skipping of alternative exons (Hnilicova et al., 2011).

H3K₂₇me₂ was not extensively investigated in the context of alternative splicing so far but have been found to be specifically enriched on exons, with the strongest presence at low expressed exons, also shown for H3K₂₇me₃ (Andersson et al., 2009; Spies et al., 2009).

The finding that most of the factors recruited by H3K₂₇me₂ are associated with the spliceosome strongly correlates this modification site with splicing activities and implicates an active functional association. Contrary to the literature, in my study an association with the splicing process was exclusively found for H3K₂₇me₂. As chromatin structure and recruitment of effector proteins by distinct histone modification patterns seem to correlate with alternative splicing it is very likely that the presence of H3K₂₇me₂ at certain chromatin regions might contribute to splice sites selection or even promote splicing events by slowing down transcriptional elongation as shown in the context of H3K9me₃.

H3K₂₇me₃ is one of the best studied chromatin modification that is catalyzed by the repressive multi-subunit complex PRC2 (Schuettengruber et al., 2007). Several polycomb group proteins, the components of the PRC2 and PRC1 complexes, have been shown to co-localize with H3K₂₇me₃ and directly targeting this modification mark (Morey and Helin, 2010; Schuettengruber et al., 2007). Surprisingly, the H3K₂₇me₃ interactome characterized here did not show an increased affinity for most of the polycomb group proteins expected to bind this mark. While I only found CBX8 and PHF1, which were significantly recruited by H3K₂₇me₃, the majority of the polycomb proteins were identified below the P-value that was set for statistical significance or they were not quantified. The mechanisms that recruit the PRC complexes to their target genes are not fully understood yet, but they are thought to contain specific and

complex elements, i.e. DNA, RNA and DNA-associated proteins (Morey and Helin, 2010). Those elements might have been absent from the chromatin template or the nuclear extracts. The preparation of nuclear extracts used for the ChAP experiments were based on high salt extraction of isolated nuclei followed by the removal of the insoluble chromatin fraction by centrifugation (Dignam et al., 1983). Thus, proteins displaying a very strong affinity to the chromatin, which resists to high salt treatment, might have been lost or reduced together with the chromatin during nuclear extract preparation.

Histone H4 methylation sites

The interactomes of H4K20me1 and H4K20me3 showed high divergence (Figure 3.9, Table 3.1) since only an overlap of three factors were found to be regulated by both modification degrees. These factors were excluded from binding to the modified chromatin templates (Figure 3.13 B).

In contrast to most of the other investigated PTMs, H4K20me1 recruited predominantly factors associated with positive regulation of transcription (Figure 3.9 B, upper panel, Table 3.1). Among them are members of the INO80 remodeling complex, which have been shown to be excluded from binding to chromatin by H3K9me1/-me3 and H3K_C27me2 histone modifications. In addition, an antagonistic regulation was observed of splicing associated factors: while H3K_C27me2 recruited several spliceosomal factors, H4K20me1 excluded 16 factors mainly associated with the C spliceosomal complex (Figure 3.9 B lower panel, Table 3.1). The data are indicative of a connection of H4K20me1 to transcriptional activity and euchromatin, which would be in agreement with previous findings that found H4K20me1 located at gene bodies (Barski et al., 2007; Vakoc et al., 2006). Additionally, H4K20me1 has been shown to be involved in regulation of many cellular processes like genome stability, DNA replication, mitosis, and transcription (Beck et al., 2012; Lv et al., 2016). The broad spectrum of biological functions linked to H4K20me1 is in good agreement with the variety of functions connected to the proteins regulated by H4K20me1 as well as the stronger differences and smaller similarities to the other histone modifications investigated here.

The second modification in this study that showed a unique protein-binding regulation when compared to the other modifications was H4R3me2. Although this modification was expected to be in association with heterochromatin (Le Guezennec et al., 2006; Zhao et al., 2009), six of nine recruited proteins were polymerase subunits. Additionally, several proteins that were recruited to H3K9me3, the hallmark of pericentric heterochromatin, were excluded from binding to H4R3me2 (Table 3.1). These findings strongly suggest an association of H4R3me2 with transcriptional activation. In agreement, it was shown that dependent on the methyltransferase

that set the di-methylation of H4R3, either PRMT1 or PRMT5, H4R3me2 can be associated with transcriptional activation or repression, respectively (Barrero and Malik, 2006; Le Guezennec et al., 2006).

4.3 Protein-binding interactomes of modified chromatin arrays display only moderate overlap compared to those revealed with different modified templates

In recent years several laboratories have made attempts to identify proteins specifically associated with distinct histone modifications by using templates such as N-terminal histone peptides, mononucleosomes and chromatin arrays (Bartke et al., 2010; Bluhm et al., 2016; Engelen et al., 2015; Kunowska et al., 2015; Nikolov et al., 2011; Oda et al., 2010; Vermeulen et al., 2010).

In the core of H3K9me3, the proteins CBX5, CBX3 and POGZ were identified in six independent studies (including this one) using different templates, cell lines and organisms (Bartke et al., 2010; Bluhm et al., 2016; Engelen et al., 2015; Kunowska et al., 2015; Nikolov et al., 2011; Vermeulen et al., 2010). This high level of convergence suggests a strong affinity to chromatin for these proteins, which is in contrast to the finding that the interaction of POGZ and CBX5 weaken the interaction of CBX5 to chromatin (Nozawa et al., 2010). Overall, nearly half of the proteins I found to be recruited by H3K9me3 were identified at least in one of the published studies mentioned above.

The overlap of H3K₂₇me3 recruited factors with published data was much lower. Common factors were CBX8 and several members of the ORC complex (Bartke et al., 2010; Vermeulen et al., 2010). The ORC complex seems to have a very strong affinity for heterochromatic modifications as the complex was also recruited to H3K9me3 and H4K20me3, which was observed using N-terminal histone peptides and mononucleosomes as well (Bartke et al., 2010; Vermeulen et al., 2010).

Next to the ORC complex I found another common feature of H3K9me3, H3K27me3 and H4K20me3 that had not been described before. The interactomes showed the recruitment of several eukaryotic translation initiation factors (Table 3.1). This was unexpected since these factors function in translation initiation within the cytoplasm and were previously shown to be absent from the nucleus (Bohnsack et al., 2002). As already discussed earlier, it is unlikely but I cannot rule out that such observation could be attributed to contaminations of the nuclear extracts by cytosolic proteins. Notably, several translation initiation factors were also found as background binders in the context of peptide pull-downs of H3K9me3, H3K27me3 and

H4K20me3 (Vermeulen et al., 2010), indicating that these factors seem to be general constituents of (HeLa) nuclear extracts. The functional association of the translation initiation factors to the heterochromatic modifications H3K9me3, H3K27me3 and H4K20me3 in contrast remains to be investigated.

Most of the differences observed are likely attributed to the different types of templates used in the previous studies. In contrast to N-terminal histone peptides, chromatin arrays are formed by an underlying DNA sequence and histone octamers. Thus, chromatin arrays allow the identification of proteins that require a binding platform consisting of more than one interaction site that can be provided by e.g. the globular domain of a histone, the DNA or a second nucleosome in the case of inter-nucleosome binding of histone modification *in trans*. The latter binding mode can also not be addressed using mononucleosomes.

Some of the differences can also likely be attributed to the different experimental procedures used in the different studies. For example, nuclear extracts derived from different organism and cell types have been used, different mass spectrometer devices with varying sensitivity, the number of technical and biological replicates, and statistical evaluations contribute to divergent protein identifications.

However, in comparison to the interactomes of defined posttranslational chromatin modifications published so far, the number of proteins we found to be affected using modified chromatin arrays is much larger. Only a moderate overlap was observed between my thesis work and the studies available in the literature (Table 3.1). This is especially true regarding proteins showing decreased affinities to a certain modification state. Thus, the data presented here offer a close to comprehensive list of factors whose binding to chromatin is affected by posttranslational modifications comprising also the excluded proteins, which in previous studies have been essentially kept out of focus. These extended dataset provide not only new functional relations of histone modifications but also enables the investigation of chromatin modification crosstalk. All these facts validate the gain of additional information provided with the here presented study and strongly support the usage of a more complex template closer to the native form of chromatin.

4.4 Individual heterochromatic marks display positive as well as negative functional correlation

Since all investigated modification states have been shown to be associated with heterochromatin and transcriptional repression (Table 1.1) I expected a certain level of overlap of regulated factors. However, due to the vast number and diversity of proteins regulated by the

individual chromatin modifications, it turned out to be very challenging to draw any general conclusions. Therefore, the functional relationship between individual modifications was investigated by using gene annotation enrichment analyses. These analyses showed positive as well as negative functional correlation of chromatin modifications and highlighted groups, which appeared to be functionally linked.

H3K9me3, H4K20me3, H3K_C27me3, and H3K9me3|H4K20me3 showed a strong correlation in the context of annotated pathway terms (Figure 3.19 A), suggesting a functional link between these modifications. All these modifications have been shown to localize to pericentric heterochromatin and to be involved in transcriptional repression (Barski et al., 2007; Fischle et al., 2003a; Lachner et al., 2003; Ringrose and Paro, 2004; Schotta et al., 2004; Sims et al., 2006; Wang et al., 2008). This observation is consistent with the strong correlation on one hand and validates the analytical approach on the other hand.

Negative functional correlation was observed between H4R3me2 and the two modification states H3K9me3 and H3K_C27me2. While H4R3me2 recruited several polymerase subunits these factors were excluded from binding to H3K9me3 and H3K_C27me2, strongly suggesting an association of H4R3me2 with transcriptional activity. H4R3me2 has been found in the context of both, transcriptional activation and repression as described earlier.

Additionally, H3K_C27me2 displayed a negative correlation to H4K20me1. While H3K_C27me2 showed an increased binding affinity for several spliceosomal factors, in turn H4K20me1 significantly excluded factors associated with the spliceosome from binding to chromatin. This is the first time that such functional relationship between both modifications is described and suggested that H4K20me1 might mark sequences at the gene bodies with decreased splicing activity, while H3K_C27me2 might promote splicing events as suggested in chapter 4.2.

Besides H3K_C27me2, H3K_C27me1, H3K9me1/-me2, and H4K20me3 also displayed negative correlation to H4K20me1 (Figure 3.19 B), which indicate a unique biological significance for H4K20me1 compared to the other investigated modification patterns.

4.5 Combinatorial readout *in trans* provides new insights into crosstalk between chemical chromatin modifications

The identification of proteins binding in the context of the combination of modifications is still challenging despite vast technical developments. Thus, the multivalent interactions of chromatin-binding proteins with combinatorial histone modification patterns were only touched upon in a limited number of studies. Here, I took the opportunity offered by our chromatin system to investigate the effect of combinatorial readout of H3K9me3 and H4K20me3 as well as H3K9me3 and CpG-methylated DNA on a global scale. Both combinations of chromatin modifications were

shown to be not only functionally linked but also displayed co-occurrence in several model organisms (Du et al., 2015; Mikkelsen et al., 2007; Schotta et al., 2004; Sims et al., 2006).

The data collected highlight three general effects of combinatorial readout that were observed for both combinations, H3K9me3|H4K20me3 and H3K9me3|meCpG.

In the first category, the double modification appeared to have “no effect” on protein-binding regulation. “No effect” was determined for the majority of proteins and referred to the finding that the protein-binding properties were similar over all three modification patterns, the two single modification sites and the combination thereof. These findings suggest that the regulation of protein recruitment to chromatin essentially relies on single modification sites.

In the second category, I observed antagonistic binding properties for several proteins when comparing the double modification to the single ones. In the context of the double modification, the decrease in binding of factors that were found otherwise to be enriched with the individual modifications was defined as negative crosstalk. In relation to H3K9me3 and H4K20me3 negative crosstalk was observed for several members of the septin family and spliceosomal factors, ASF1B, EIF5 and EIF4G1 (Figure 3.16 A). These findings give insights into the extent of consequences brought about by crosstalk. The effects of positive and negative crosstalk were not only limited to individual proteins but also were applied to entire protein complexes. Thus, it seems that not only direct binding proteins but also secondary interacting proteins were affected by crosstalk. Additionally, individual factors displayed negative crosstalk between H3K9me3 and meCpG as well as H3K9me3 and H4K20me3, namely USP3 and UBTF (Figure 3.16 A and 3.18 A).

Opposed binding was also determined in the reverse direction: crosstalk was considered as positive when the double modification displayed increased protein binding, while the individual modifications displayed decreased binding to chromatin. Several polymerase subunits and members of the NFI family displayed positive crosstalk between both H3K9me3 and meCpG as well as H3K9me3 and H4K20me3. Additionally, each of the modification combinations showed individual transcription factors that were affected. Recently, positive crosstalk was shown in context of H3 methylation and H4 acetylation. It was shown that the presence of H4K5ac next to H3K4me3 enhances the binding of BPTF and p300 to nucleosomes by at least seven-fold, probably mediated by simultaneous binding of the PHD and the bromodomain (Nguyen et al., 2014).

In the present study, conclusions about the binding mechanisms of single factors could not be drawn, as the experimental design does not allow determining whether a given protein binds directly or indirectly to the chromatin arrays. However, it appeared that positive crosstalk of chromatin modifications, mainly located at pericentric heterochromatin, tends to recruit factors

and protein complexes in association with transcriptional activity and thus, might be involved in the reversion of transcriptional repression at defined chromatin regions.

Proteins showing negative crosstalk are inconclusively transcriptional repressors or in association with heterochromatin. In the context of negative crosstalk there seemed to be a trend of excluding factors or protein complexes involved in mitosis and proteins associated with the cytoskeleton in presence of one of the here investigated double modifications H3K9me3|meCpG and H3K9me3|H4K20me3.

4.6 Protein-protein cross-linking coupled to MS maps the binding hierarchy of chromatin-bound proteins

So far the identification of specific protein interactions using affinity purification based strategies coupled with cross-linking and mass spectrometry were challenging, mainly because of the lack in database engines capable to search for cross-links against large databases, which has restricted cross-linking approaches to the study of small protein complexes for a long time.

Now, with important technological developments of recent years, it has become possible to set up a workflow that combines proteomic and protein-specific XL-MS. This offers the possibility to analyze topological information of the protein interactome associated with chromatin. As I have demonstrated, both the protein-binding interactome as well as the hierarchy of protein binding to chromatin can be analyzed.

With the established workflow I found 75 cross-linked chromatin-associated proteins. As cross-links covalently link proteins that are in close physical proximity, the proteins directly cross-linked to histones were considered as primary interactors. 18 of the cross-linked proteins showed a direct physical interaction to at least one of the core histones (Figure 3.26 A) while 20 factors were identified as indirect binders, not cross-linked to histones, but recruited to the chromatin arrays via primary binders. The remaining identified proteins could not be assigned to one of the core histones. The connecting factors may have not been detected or, alternatively, these proteins are recruited to chromatin by interacting with the DNA.

STRING analysis of cross-linked proteins confirmed the biological association of these proteins and came as a support regarding the specificity of the cross-links and of the approach on a more general way (Figure 3.26 B).

In contrast to the fact that more than 1000 proteins were identified to bind to the chromatin arrays using ChAP-MS, only a small number of proteins were found to be cross-linked. This low number might be attributed to limited availability and accessibility of interaction mediating sequences qualified for cross-linking. In brief, the BS3 cross-linker, which comprises two amine

reactive groups separated by a spacer arm of 11.4-Å in length, reacts with primary amines in the side chains of lysine (K) residues and the N-terminus of polypeptides (Sinz, 2006; Tran et al., 2016). Thus, only interacting protein surfaces exhibiting two lysines within a distance of 24-Å can be cross-linked. As a consequence, interaction interface that do not have lysines matching this criteria cannot be cross-linked, which result in the impossibility to identify actual interactions. This limitation may account, at least in part, for the rather small number of cross-linked proteins which were identified. Moreover, BS3 tend mostly to react with easily accessible lysine residues localized at the surfaces of protein complexes, since diffusion within the rather hydrophobic globular part of complexes is hindered by its hydrophilic nature (Huang et al., 2004). As a consequence, this may lower the number of cross-links and therefore reduce the number of final interacting partner identified. Additionally, the two enrichment steps on the protein as well as the peptide level by size exclusion chromatography are known to lead to loss of biological material.

Another limitation is related to on-bead cross-linking. Streptavidin coated beads that were used in this workflow have the potential to cross-link purified proteins to the beads itself, potentially leading to false positive identifications. To overcome this issue, the underlying biotinylated DNA of the cross-linked sample was degraded, which resulted in the release of histone octamers and their associated proteins from the beads.

In conclusion, we established a robust workflow that allows the highly specific identification of protein-protein cross-links in the context of chromatin affinity purification. The workflow enables the mapping of specific protein-protein interaction sites on one hand and provides information regarding the hierarchy of protein binding with each other and to chromatin on the other hand. The workflow identifies the abundant chromatin-interacting proteins and therefore has a strong potential for the investigation of physical protein-protein interaction sites of protein interactomes of modified chromatin arrays.

5 Conclusions and future perspectives

I have shown here, that ChAP-MS is suitable for the elucidation of the interactomes of single and complex chromatin modification patterns. My results provide a step towards functional characterization of individual modifications deposited on an oligonucleosomal chromatin template. Each of the modification patterns studied here regulated a specific set of factors, underscoring their distinct biological significance. The protein interactomes of individual modifications have revealed novel functional associations of specific marks. This was shown for H3K₂₇me₂ that recruits several splicing factors, H4R3me₂, which appears to have mainly impact on the recruitment of polymerase subunits and H3K₂₇me₁, which was associated with transcriptional initiation and elongation by the CCR4-NOT complex. The assignment of biological functions to the different chromatin-binding interactomes on a global scale allowed the identification of novel relationships between individual modifications and indicated positive and negative functional correlation of PTMs. One example is the inverse regulation of the binding properties of several spliceosomal factors by H3K₂₇me₂ and H4K20me₁.

My work shows that combinations of different modifications in the context of chromatin can reveal important aspects of the nature and the complexity of the language of histone PTMs. Indeed, the combinatorial experiments presented here identified a subset of factors regulated only in presence of two modification sites. Thus, my experimental design gives novel, unpredicted insights into mechanisms of crosstalk between two chromatin modifications. Both, H3K9me₃ and mCpG as well as H3K9me₃ and H4K20me₃ display positive as well as negative crosstalk, which underlines the importance of testing further combinations of modifications whose co-occurrence have been proved *in vivo* already.

The performed experiments have proven to provide new insights into specific functions and regulation mechanisms of individual histone modification patterns. Extending this approach to euchromatic marks would significantly increase the knowledge of pathways underlying epigenetic regulations. Furthermore, ChAP-MS can determine the protein-binding interactomes in the context of distinct biological circumstances by using e.g. a native DNA sequence instead of an artificial one or cell extracts from different cell types, organisms or different developmental stages.

The new cross-linking approach we have established provides a workflow for the identification of specific protein-protein interaction of chromatin-bound proteins. I expect this method will contribute to our understanding of the hierarchy of protein recruitment and deliver detailed information of physical interactions sites, allowing the identification of protein domains involved in protein-protein interactions. Expanding the workflow to modified chromatin arrays will in the near future provide insights into the formation of protein complexes in the context of different chromatin modification patterns.

6 References

- Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol Cell* 5, 367-375.
- Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., and Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome research* 19, 1732-1741.
- Angelov, D., Vitolo, J.M., Mutskov, V., Dimitrov, S., and Hayes, J.J. (2001). Preferential interaction of the core histone tail domains with linker DNA. *Proceedings of the National Academy of Sciences of the United States of America* 98, 6599-6604.
- Arita, K., Isogai, S., Oda, T., Unoki, M., Sugita, K., Sekiyama, N., Kuwata, K., Hamamoto, R., Tochio, H., Sato, M., *et al.* (2012). Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *Proceedings of the National Academy of Sciences of the United States of America* 109, 12950-12955.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* 21, 381-395.
- Bantscheff, M., Lemeer, S., Savitski, M.M., and Kuster, B. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* 404, 939-965.
- Barrero, M.J., and Malik, S. (2006). Two functional modes of a nuclear receptor-recruited arginine methyltransferase in transcriptional activation. *Mol Cell* 24, 233-243.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823-837.
- Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* 143, 470-484.
- Beck, D.B., Oda, H., Shen, S.S., and Reinberg, D. (2012). PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes & development* 26, 325-337.
- Berger, S.L., Kouzarides, T., Shiekhata, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes & development* 23, 781-783.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.
- Blencowe, B.J., Issner, R., Nickerson, J.A., and Sharp, P.A. (1998). A coactivator of pre-mRNA splicing. *Genes & development* 12, 996-1009.

- Bluhm, A., Casas-Vila, N., Scheibe, M., and Butter, F. (2016). Reader interactome of epigenetic histone marks in birds. *Proteomics* 16, 427-436.
- Blum H., B.H., Gross h.J. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8, 93-99.
- Bohnsack, M.T., Regener, K., Schwappach, B., Saffrich, R., Paraskeva, E., Hartmann, E., and Gorlich, D. (2002). Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. *The EMBO journal* 21, 6205-6215.
- Bonasio, R., Lecona, E., and Reinberg, D. (2010). MBT domain proteins in development and disease. *Semin Cell Dev Biol* 21, 221-230.
- Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760-1764.
- Brasher, S.V., Smith, B.O., Fogh, R.H., Nietlispach, D., Thiru, A., Nielsen, P.R., Broadhurst, R.W., Ball, L.J., Murzina, N.V., and Laue, E.D. (2000). The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *The EMBO journal* 19, 1587-1597.
- Breiling, A., and Lyko, F. (2015). Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* 8, 24.
- Bronner, C., Achour, M., Arima, Y., Chataigneau, T., Saya, H., and Schini-Kerth, V.B. (2007). The UHRF family: oncogenes that are drugable targets for cancer therapy in the near future? *Pharmacol Ther* 115, 419-434.
- Chan, D.W., Wang, Y., Wu, M., Wong, J., Qin, J., and Zhao, Y. (2009). Unbiased proteomic screen for binding proteins to modified lysines on histone H3. *Proteomics* 9, 2343-2354.
- Chen, X., Wei, S., Ji, Y., Guo, X., and Yang, F. (2015). Quantitative proteomics using SILAC: Principles, applications, and developments. *Proteomics* 15, 3175-3192.
- Collart, M.A. (2016). The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley interdisciplinary reviews RNA*.
- Combe, C.W., Fischer, L., and Rappsilber, J. (2015). xiNET: cross-link network maps with residue resolution. *Molecular & cellular proteomics : MCP* 14, 1137-1147.
- Cowieson, N.P., Partridge, J.F., Allshire, R.C., and McLaughlin, P.J. (2000). Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Curr Biol* 10, 517-525.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 26, 1367-1372.
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* 10, 1794-1805.

- Cui, X.J., and Shi, C.X. (2016). Combinations of Histone Modifications for Pattern Genes. *Acta Biotheor.*
- Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *Journal of molecular biology* 319, 1097-1113.
- Demeler, B. UltraScan (current version), A comprehensive software package for the analysis of sedimentation experiments. Dept of Biochemistry, The University of Texas Health Science Center.
- Di Palma, S., Hennrich, M.L., Heck, A.J., and Mohammed, S. (2012). Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J Proteomics* 75, 3791-3813.
- Dignam, J.D., Martin, P.L., Shastry, B.S., and Roeder, R.G. (1983). Eukaryotic gene transcription with purified components. *Methods in enzymology* 101, 582-598.
- Dillon, N., and Festenstein, R. (2002). Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet* 18, 252-258.
- Dimitrov, A., Paupe, V., Gueudry, C., Sibarita, J.B., Raposo, G., Vilemeyer, O., Gilbert, T., Csaba, Z., Attie-Bitach, T., Cormier-Daire, V., *et al.* (2009). The gene responsible for Dyggve-Melchior-Clausen syndrome encodes a novel peripheral membrane protein dynamically associated with the Golgi apparatus. *Hum Mol Genet* 18, 440-453.
- Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* 306, 1571-1573.
- Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* 16, 519-532.
- Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods in enzymology* 375, 23-44.
- Elgin, S.C., and Grewal, S.I. (2003). Heterochromatin: silence is golden. *Curr Biol* 13, R895-898.
- Engelen, E., Brandsma, J.H., Moen, M.J., Signorile, L., Dekkers, D.H., Demmers, J., Kockx, C.E., Ozgur, Z., van, I.W.F., van den Berg, D.L., *et al.* (2015). Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat Commun* 6, 7155.
- Fairley, J.A., Scott, P.H., and White, R.J. (2003). TFIIIB is phosphorylated, disrupted and selectively released from tRNA promoters during mitosis in vivo. *The EMBO journal* 22, 5841-5850.
- Ferri, E., Petosa, C., and McKenna, C.E. (2016). Bromodomains: Structure, function and pharmacology of inhibition. *Biochem Pharmacol* 106, 1-18.

- Fischle, W., Wang, Y., and Allis, C.D. (2003a). Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475-479.
- Fischle, W., Wang, Y., and Allis, C.D. (2003b). Histone and chromatin cross-talk. *Curr Opin Cell Biol* **15**, 172-183.
- Fritzsche, R., Ihling, C.H., Gotze, M., and Sinz, A. (2012). Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. *Rapid Commun Mass Spectrom* **26**, 653-658.
- Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic acids research* **31**, 2305-2312.
- Gallagher, S.R. (2006). One-dimensional SDS gel electrophoresis of proteins. *Current protocols in molecular biology* / edited by Frederick M Ausubel [et al] *Chapter 10*, Unit 10 12A.
- Gayatri, S., and Bedford, M.T. (2014). Readers of histone methylarginine marks. *Biochimica et biophysica acta* **1839**, 702-710.
- Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R., and Mann, M. (2010). Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* **7**, 383-385.
- Geiger, T., Wisniewski, J.R., Cox, J., Zanivan, S., Kruger, M., Ishihama, Y., and Mann, M. (2011). Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature protocols* **6**, 147-157.
- Glover, J.N. (2006). Insights into the molecular basis of human hereditary breast cancer from studies of the BRCA1 BRCT domain. *Fam Cancer* **5**, 89-93.
- Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* **74**, 481-514.
- Gromak, N., Rideau, A., Southby, J., Scadden, A.D., Gooding, C., Huttelmaier, S., Singer, R.H., and Smith, C.W. (2003). The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing. *The EMBO journal* **22**, 6356-6364.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88.
- Heemskerk, A.A., Deelder, A.M., and Mayboroda, O.A. (2016). CE-ESI-MS for bottom-up proteomics: Advances in separation, interfacing and applications. *Mass spectrometry reviews* **35**, 259-271.
- Heitz, E. (1928). Das Heterochromatin der Moose. *I Jahrb Wiss Bot* **69**, 762-819.
- Hnilicova, J., Hozeifi, S., Duskova, E., Icha, J., Tomankova, T., and Stanek, D. (2011). Histone deacetylase activity modulates alternative splicing. *PloS one* **6**, e16727.
- Hou, Z., Peng, H., Ayyanathan, K., Yan, K.P., Langer, E.M., Longmore, G.D., and Rauscher, F.J., 3rd (2008). The LIM protein AJUBA recruits protein arginine methyltransferase 5 to mediate SNAIL-dependent transcriptional repression. *Molecular and cellular biology* **28**, 3198-3207.

- Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., *et al.* (2013). DNA methylation presents distinct binding sites for human transcription factors. *Elife* 2, e00726.
- Huang, B.X., Kim, H.Y., and Dass, C. (2004). Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *Journal of the American Society for Mass Spectrometry* 15, 1237-1247.
- Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37, 1-13.
- Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4, 44-57.
- Huynh, V.A., Robinson, P.J., and Rhodes, D. (2005). A method for the in vitro reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone. *Journal of molecular biology* 345, 957-968.
- Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. *FEBS Lett* 583, 1713-1720.
- Izzo, A., and Schneider, R. (2010). Chatting histone modifications in mammals. *Brief Funct Genomics* 9, 429-443.
- Jacob, Y., Stroud, H., Leblanc, C., Feng, S., Zhuo, L., Caro, E., Hassel, C., Gutierrez, C., Michaels, S.D., and Jacobsen, S.E. (2010). Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 466, 987-991.
- Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* 295, 2080-2083.
- Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., *et al.* (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research* 37, D412-416.
- Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074-1080.
- Jin, B., Ernst, J., Tiedemann, R.L., Xu, H., Sureshchandra, S., Kellis, M., Dalton, S., Liu, C., Choi, J.H., and Robertson, K.D. (2012). Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. *Cell Rep* 2, 1411-1424.
- Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13, 484-492.
- Karagianni, P., Amazit, L., Qin, J., and Wong, J. (2008). ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Molecular and cellular biology* 28, 705-717.
- Kim, D., Blus, B.J., Chandra, V., Huang, P., Rastinejad, F., and Khorasanizadeh, S. (2010). Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nature structural & molecular biology* 17, 1027-1029.

- Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* 7, 397-403.
- Kimura, H. (2013). Histone modifications for human epigenome analysis. *J Hum Genet* 58, 439-445.
- Kokura, K., Sun, L., Bedford, M.T., and Fang, J. (2010). Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. *The EMBO journal* 29, 3673-3687.
- Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* 184, 868-871.
- Kosinski, J., von Appen, A., Ori, A., Karius, K., Muller, C.W., and Beck, M. (2015). Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *Journal of structural biology* 189, 177-183.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.
- Kruithof, M., Chien, F.T., Routh, A., Logie, C., Rhodes, D., and van Noort, J. (2009). Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nature structural & molecular biology* 16, 534-540.
- Kudo, N., Khochbin, S., Nishi, K., Kitano, K., Yanagida, M., Yoshida, M., and Horinouchi, S. (1997). Molecular cloning and cell cycle-dependent expression of mammalian CRM1, a protein involved in nuclear export of proteins. *The Journal of biological chemistry* 272, 29742-29751.
- Kunowska, N., Rotival, M., Yu, L., Choudhary, J., and Dillon, N. (2015). Identification of protein complexes that bind to histone H3 combinatorial modifications using super-SILAC and weighted correlation network analysis. *Nucleic acids research* 43, 1418-1432.
- Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 484, 115-119.
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116-120.
- Lachner, M., O'Sullivan, R.J., and Jenuwein, T. (2003). An epigenetic road map for histone lysine methylation. *Journal of cell science* 116, 2117-2124.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Langelier, M.F., Planck, J.L., Roy, S., and Pascal, J.M. (2012). Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. *Science* 336, 728-732.
- Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet* 32, 42-56.

- Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Molecular and cellular biology* 26, 843-851.
- Lee, J.S., Smith, E., and Shilatifard, A. (2010). The language of histone crosstalk. *Cell* 142, 682-685.
- Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* 13, 1192-1200.
- Leitner, A., Reischl, R., Walzthoeni, T., Herzog, F., Bohn, S., Forster, F., and Aebersold, R. (2012). Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Molecular & cellular proteomics : MCP* 11, M111 014126.
- LeRoy, G., Orphanides, G., Lane, W.S., and Reinberg, D. (1998). Requirement of RSF and FACT for transcription of chromatin templates in vitro. *Science* 282, 1900-1904.
- Li, Z., Dai, H., Martos, S.N., Xu, B., Gao, Y., Li, T., Zhu, G., Schones, D.E., and Wang, Z. (2015). Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. *Genome biology* 16, 115.
- Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C., *et al.* (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature genetics* 47, 469-478.
- Licklider, L.J., Thoreen, C.C., Peng, J., and Gygi, S.P. (2002). Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Analytical chemistry* 74, 3076-3083.
- Lis, J.T., and Schleif, R. (1975). Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. *Nucleic acids research* 2, 383-389.
- Litt, M., Qiu, Y., and Huang, S. (2009). Histone arginine methylations: their roles in chromatin dynamics and transcriptional regulation. *Biosci Rep* 29, 131-141.
- Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun* 4, 1563.
- Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J.P., Imhof, A., Nakatani, Y., Dent, S.Y., and Almouzni, G. (2009). The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Rep* 10, 769-775.
- Lu, R., and Wang, G.G. (2013). Tudor: a versatile family of histone methylation 'readers'. *Trends Biochem Sci* 38, 546-555.
- Lu, X., Simon, M.D., Chodaparambil, J.V., Hansen, J.C., Shokat, K.M., and Luger, K. (2008). The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nature structural & molecular biology* 15, 1122-1124.

- Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell* **144**, 16-26.
- Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* **327**, 996-1000.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260.
- Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. *Methods in enzymology* **304**, 3-19.
- Lv, X., Han, Z., Chen, H., Yang, B., Yang, X., Xia, Y., Pan, C., Fu, L., Zhang, S., Han, H., *et al.* (2016). A positive role for polycomb in transcriptional regulation via H4K20me1. *Cell Res.*
- Makiniemi, M., Hillukkala, T., Tuusa, J., Reini, K., Vaara, M., Huang, D., Pospiech, H., Majuri, I., Westerling, T., Makela, T.P., *et al.* (2001). BRCT domain-containing protein TopBP1 functions in DNA replication and damage response. *The Journal of biological chemistry* **276**, 30399-30406.
- Margueron, R., Trojer, P., and Reinberg, D. (2005). The key to development: interpreting the histone code? *Curr Opin Genet Dev* **15**, 163-176.
- Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466**, 1120-1124.
- Marton, M.J., Vazquez de Aldana, C.R., Qiu, H., Chakraborty, K., and Hinnebusch, A.G. (1997). Evidence that GCN1 and GCN20, translational regulators of GCN4, function on elongating ribosomes in activation of eIF2α kinase GCN2. *Molecular and cellular biology* **17**, 4474-4489.
- Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* **28**, 69-74.
- McBryant, S.J., and Hansen, J.C. (2012). Dynamic fuzziness during linker histone action. *Adv Exp Med Biol* **725**, 15-26.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560.
- Montini, E., Buchner, G., Spalluto, C., Andolfi, G., Caruso, A., den Dunnen, J.T., Trump, D., Rocchi, M., Ballabio, A., and Franco, B. (1999). Identification of SCML2, a second human gene homologous to the Drosophila sex comb on midleg (Scm): A new gene cluster on Xp22. *Genomics* **58**, 65-72.
- Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci* **35**, 323-332.
- Mosch, K., Franz, H., Soeroes, S., Singh, P.B., and Fischle, W. (2011). HP1 recruits activity-dependent neuroprotective protein to H3K9me3 marked pericentromeric heterochromatin for silencing of major satellite repeats. *PloS one* **6**, e15894.

Musselman, C.A., and Kutateladze, T.G. (2011). Handpicking epigenetic marks with PHD fingers. *Nucleic acids research* 39, 9061-9071.

Musselman, C.A., Lalonde, M.E., Cote, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. *Nature structural & molecular biology* 19, 1218-1227.

Mutskov, V., Gerber, D., Angelov, D., Ausio, J., Workman, J., and Dimitrov, S. (1998). Persistent interactions of core histone tails with nucleosomal DNA following acetylation and transcription factor binding. *Molecular and cellular biology* 18, 6293-6304.

Nady, N., Lemak, A., Walker, J.R., Avvakumov, G.V., Kareta, M.S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X., *et al.* (2011). Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *The Journal of biological chemistry* 286, 24300-24311.

Nan, X., Meehan, R.R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic acids research* 21, 4886-4892.

Nguyen, U.T., Bittova, L., Muller, M.M., Fierz, B., David, Y., Houck-Loomis, B., Feng, V., Dann, G.P., and Muir, T.W. (2014). Accelerated chromatin biochemistry using DNA-barcoded nucleosome libraries. *Nat Methods* 11, 834-840.

Nikolov, M., Stutzer, A., Mosch, K., Krasauskas, A., Soeroes, S., Stark, H., Urlaub, H., and Fischle, W. (2011). Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns. *Molecular & cellular proteomics : MCP* 10, M110 005371.

Nozawa, R.S., Nagao, K., Masuda, H.T., Iwasaki, O., Hirota, T., Nozaki, N., Kimura, H., and Obuse, C. (2010). Human POGZ modulates dissociation of HP1alpha from mitotic chromosome arms through Aurora B activation. *Nat Cell Biol* 12, 719-727.

Oda, H., Hubner, M.R., Beck, D.B., Vermeulen, M., Hurwitz, J., Spector, D.L., and Reinberg, D. (2010). Regulation of the histone H4 monomethylase PR-Set7 by CRL4(Cdt2)-mediated PCNA-dependent degradation during DNA damage. *Mol Cell* 40, 364-376.

Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* 105, 487-497.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). *Science* 183, 330-332.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* 1, 376-386.

Ong, S.E., and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 1, 252-262.

Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *The EMBO journal* 19, 6141-6149.

- Paramelle, D., Miralles, G., Subra, G., and Martinez, J. (2013). Chemical cross-linkers for protein structure studies by mass spectrometry. *Proteomics* 13, 438-456.
- Patel, D.J. (2016). A Structural Perspective on Readout of Epigenetic Histone and DNA Methylation Marks. *Cold Spring Harb Perspect Biol* 8.
- Peng, W., Togawa, C., Zhang, K., and Kurdistani, S.K. (2008). Regulators of cellular levels of histone acetylation in *Saccharomyces cerevisiae*. *Genetics* 179, 277-289.
- Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., *et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* 12, 1577-1589.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605-1612.
- Plafker, S.M., and Macara, I.G. (2000). Importin-11, a nuclear import receptor for the ubiquitin-conjugating enzyme, UbcM2. *The EMBO journal* 19, 5502-5513.
- Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & development* 15, 1613-1618.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 81, 559-575.
- R Development-Core-Team. (2011). R: A Language and Environment for Statistical Computing. (Vienna, Austria: R Foundation for Statistical Computing).
- Ramon-Maiques, S., Kuo, A.J., Carney, D., Matthews, A.G., Oettinger, M.A., Gozani, O., and Yang, W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proceedings of the National Academy of Sciences of the United States of America* 104, 18993-18998.
- Reddy, J.V., Burguete, A.S., Sridevi, K., Ganley, I.G., Nottingham, R.M., and Pfeffer, S.R. (2006). A functional role for the GCC185 golgin in mannose 6-phosphate receptor recycling. *Molecular biology of the cell* 17, 4353-4363.
- Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annual review of genetics* 38, 413-443.
- Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L.N., Beck, M., Schmidt, A., Mueller, M., and Aebersold, R. (2008). Identification of cross-linked peptides from large sequence databases. *Nat Methods* 5, 315-318.
- Rosenfeld, J.A., Wang, Z., Schones, D.E., Zhao, K., DeSalle, R., and Zhang, M.Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics* 10, 143.
- Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. *Biochimica et biophysica acta* 1839, 627-643.

- Ruthenburg, A.J., Li, H., Milne, T.A., Dewell, S., McGinty, R.K., Yuen, M., Ueberheide, B., Dou, Y., Muir, T.W., Patel, D.J., *et al.* (2011). Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* 145, 692-706.
- Saint-Andre, V., Batsche, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. *Nature structural & molecular biology* 18, 337-344.
- Sambrook, J., Russell, D. (2001). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press).
- Sanchez, R., Meslamani, J., and Zhou, M.M. (2014). The bromodomain: from epigenome reader to druggable target. *Biochimica et biophysica acta* 1839, 676-685.
- Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436, 138-141.
- Schmidt, C. (2010). Absolute and relative quantification of proteins in large protein-RNA assemblies by mass spectrometry. In *Bioanalytical Mass Spectrometry Group* (Göttingen: Georg August University).
- Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & development* 18, 1251-1262.
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* 128, 735-745.
- Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nature structural & molecular biology* 16, 990-995.
- Shamay, M., Barak, O., Doitsh, G., Ben-Dor, I., and Shaul, Y. (2002). Hepatitis B virus pX interacts with HBXAP, a PHD finger protein to coactivate transcription. *The Journal of biological chemistry* 277, 9982-9988.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13, 2498-2504.
- Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., *et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908-912.
- Sharma, S., Wongpalee, S.P., Vashisht, A., Wohlschlegel, J.A., and Black, D.L. (2014). Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes & development* 28, 2518-2531.
- Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols* 1, 2856-2860.
- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical chemistry* 68, 850-858.

- Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844-847.
- Shogren-Knaak, M.A., and Peterson, C.L. (2004). Creating designer histones by native chemical ligation. *Methods in enzymology* 375, 62-76.
- Simon, M.D. (2010). Installation of site-specific methylation into histones using methyl lysine analogs. *Current protocols in molecular biology* / edited by Frederick M Ausubel [et al] *Chapter* 21, Unit 21 18 21-10.
- Simon, M.D., Chu, F., Racki, L.R., de la Cruz, C.C., Burlingame, A.L., Panning, B., Narlikar, G.J., and Shokat, K.M. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* 128, 1003-1012.
- Sims, J.K., Houston, S.I., Magazinnik, T., and Rice, J.C. (2006). A trans-tail histone code defined by monomethylated H4 Lys-20 and H3 Lys-9 demarcates distinct regions of silent chromatin. *The Journal of biological chemistry* 281, 12760-12766.
- Sinz, A. (2006). Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass spectrometry reviews* 25, 663-682.
- Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* 36, 245-254.
- Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* 403, 41-45.
- Surka, M.C., Tsang, C.W., and Trimble, W.S. (2002). The mammalian septin MSF localizes with microtubules and is required for completion of cytokinesis. *Molecular biology of the cell* 13, 3532-3545.
- Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., and Guigo, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nature structural & molecular biology* 16, 996-1001.
- Torres, I.O., and Fujimori, D.G. (2015). Functional coupling between writers, erasers and readers of histone and DNA methylation. *Curr Opin Struct Biol* 35, 68-75.
- Towbin, B.D., Gonzalez-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., Askjaer, P., and Gasser, S.M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* 150, 934-947.
- Tran, B.Q., Goodlett, D.R., and Goo, Y.A. (2016). Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. *Biochimica et biophysica acta* 1864, 123-129.
- Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* 128, 651-654.
- Tsai, W.W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.Y., Shi, X., Schwarzer, D., Plunkett, W., et al. (2010). TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 468, 927-932.

- Ucar, D., Hu, Q., and Tan, K. (2011). Combinatorial chromatin modification patterns in the human genome revealed by subspace clustering. *Nucleic acids research* 39, 4063-4075.
- Unoki, M., Nishidate, T., and Nakamura, Y. (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene* 23, 7601-7610.
- Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. *Molecular and cellular biology* 26, 9185-9195.
- Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J.E., and Corces, V.G. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome research* 22, 2176-2187.
- van Kruijsbergen, I., Hontelez, S., and Veenstra, G.J. (2015). Recruiting polycomb to chromatin. *Int J Biochem Cell Biol* 67, 177-187.
- Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent histone modifications in early embryogenesis. *Curr Opin Cell Biol* 24, 374-386.
- Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., *et al.* (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 142, 967-980.
- Vezzoli, A., Bonadies, N., Allen, M.D., Freund, S.M., Santiveri, C.M., Kvinlaug, B.T., Huntly, B.J., Gottgens, B., and Bycroft, M. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nature structural & molecular biology* 17, 617-619.
- Waddington, C.H. (1953). Epigenetics and evolution. *Symposia of the Society for Experimental Biology* 7, 186-199.
- Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., Bohn, S., Forster, F., Beck, M., and Aebersold, R. (2012). False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat Methods* 9, 901-903.
- Wang, X., Takenaka, K., and Takeda, S. (2010). PTIP promotes DNA double-strand break repair through homologous recombination. *Genes Cells* 15, 243-254.
- Wang, Z., Schones, D.E., and Zhao, K. (2009). Characterization of human epigenomes. *Curr Opin Genet Dev* 19, 127-134.
- Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., *et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* 40, 897-903.
- Wendorff, T.J., Schmidt, B.H., Heslop, P., Austin, C.A., and Berger, J.M. (2012). The structure of DNA-bound human topoisomerase II alpha: conformational mechanisms for coordinating inter-subunit interactions with DNA cleavage. *Journal of molecular biology* 424, 109-124.
- Woodcock, C.L., Skoultchi, A.I., and Fan, Y. (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res* 14, 17-25.

- Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* 329, 444-448.
- Wysocka, J. (2006). Identifying novel proteins recognizing histone modifications using peptide pull-down assay. *Methods* 40, 339-343.
- Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., *et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442, 86-90.
- Xi, Q., Wang, Z., Zaromytidou, A.I., Zhang, X.H., Chow-Tsang, L.F., Liu, J.X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., *et al.* (2011). A poised chromatin platform for TGF-beta access to master regulators. *Cell* 147, 1511-1524.
- Xie, H., Surka, M., Howard, J., and Trimble, W.S. (1999). Characterization of the mammalian septin H5: distinct patterns of cytoskeletal and membrane association from other septin proteins. *Cell motility and the cytoskeleton* 43, 52-62.
- Xu, X., Hoang, S., Mayo, M.W., and Bekiranov, S. (2010). Application of machine learning methods to histone methylation ChIP-Seq data reveals H4R3me2 globally represses gene expression. *BMC Bioinformatics* 11, 396.
- Yang, N., and Xu, R.M. (2013). Structure and function of the BAH domain in chromatin biology. *Crit Rev Biochem Mol Biol* 48, 211-221.
- Yap, K.L., and Zhou, M.M. (2010). Keeping it in the family: diverse histone recognition by conserved structural folds. *Crit Rev Biochem Mol Biol* 45, 488-505.
- Yates, J.R., 3rd (2004). Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct* 33, 297-316.
- Yu, Y.Q., Gilar, M., Lee, P.J., Bouvier, E.S., and Gebler, J.C. (2003). Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Analytical chemistry* 75, 6023-6028.
- Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* 466, 258-262.
- Zhao, Q., Rank, G., Tan, Y.T., Li, H., Moritz, R.L., Simpson, R.J., Cerruti, L., Curtis, D.J., Patel, D.J., Allis, C.D., *et al.* (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nature structural & molecular biology* 16, 304-311.
- Zhou, H.L., Luo, G., Wise, J.A., and Lou, H. (2014). Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic acids research* 42, 701-713.

7 Appendix

Appendix 1 Proteins significantly recruited or excluded by one of the chromatin modification patterns investigated by chromatin affinity purification. The gene names, the uniprot identifier (Protein IDs) and the normalized heavy to light ratios of the forward (F) and the reverse (R) experiments are shown as reported by MaxQuant for each chromatin modification state.

[illegible]

Gene names	Protein IDs	H3Δ1-20_F	H3Δ1-20_R	H4K20me1_F	H4K20me1_R	H4K20me3_F	H4K20me3_R	H3Kc27me1_F	H3Kc27me1_R	H3Kc27me2_F	H3Kc27me2_R	H3Kc27me3_F	H3Kc27me3_R	H3K9me1_F	H3K9me1_R	H3K9me2_F	H3K9me2_R	H3K9me3_F	H3K9me3_R	H3K9me3 meCpG_F	H3K9me3 meCpG_R	H3K9me3 H4K20me3_F	H3K9me3 H4K20me3_R	meCpG_F	meCpG_R	H4R3me2_F	H4R3me2_R	
ATF2	P15336-5	0.28	0.37	NaN	NaN	0.49	0.94	0.63	0.59	0.56	1.10	0.71	0.64	NaN	NaN	0.06	NaN	0.39	1.13	3.35	0.21	1.08	0.37	0.83	1.02	NaN	NaN	
ATF2	P15336-8	NaN	NaN	NaN	NaN	0.46	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
ATL3	Q6DD88	NaN	0.24	0.60	NaN	NaN	0.29	NaN	NaN	0.81	NaN	NaN	NaN	NaN	NaN	NaN	2.15	2.21	NaN	0.94	NaN	NaN	1.05	NaN	NaN	NaN	NaN	
ATP6V0D1	B2R7M1	1.15	1.51	NaN	NaN	NaN	1.02	NaN	NaN	NaN	NaN	NaN	NaN	1.04	0.99	0.78	0.18	NaN	NaN	NaN	NaN	NaN	NaN	0.13	NaN	NaN	NaN	
ATP6V1B2	P21281	0.97	0.80	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.82	0.62	0.78	0.38	NaN	NaN	NaN	NaN	NaN	NaN	0.12	NaN	NaN	NaN	
ATP6V1F	A4D1K0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.21	NaN	NaN	NaN	
ATRIP	Q8WXE1-2	0.85	0.81	0.32	NaN	0.70	0.66	NaN	NaN	0.84	NaN	0.96	NaN	0.84	0.45	0.62	0.36	0.81	0.79	NaN	0.38	0.46	0.61	NaN	NaN	1.67	NaN	
ATRX	A4LAA3	0.65	0.48	0.33	0.55	0.39	0.40	0.34	0.49	0.48	0.50	0.44	0.55	0.51	1.07	0.67	0.35	1.41	0.20	1.25	0.13	0.85	0.25	0.50	0.32	0.45	0.39	
BABAM1	M0R0I0	0.84	0.40	NaN	3.17	NaN	NaN	NaN	NaN	1.34	0.55	NaN	NaN	NaN	0.89	NaN	0.25	NaN	NaN	0.63	4.44	0.23	2.81	NaN	NaN	NaN	NaN	
BACH1	Q6ICU0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.74	0.17	0.94	NaN	3.68	0.37	NaN	NaN	
BAG6	B0UX83	1.79	0.50	0.61	1.03	0.99	0.31	1.27	0.32	0.84	0.72	0.72	0.40	2.43	0.95	1.53	0.21	1.34	0.52	NaN	NaN	0.93	1.07	NaN	NaN	1.52	0.40	
BARD1	F6MDH7	0.64	0.37	0.11	4.55	0.23	2.99	0.75	0.75	1.02	0.67	0.97	0.77	0.60	1.09	0.59	1.17	0.74	0.84	0.63	1.22	0.19	2.47	1.04	0.64	0.64	0.92	
BAZ1B	Q9UIG0-2	0.07	5.41	1.21	0.93	1.31	1.00	0.98	0.86	0.93	1.19	1.04	1.10	1.08	1.38	0.67	1.02	1.10	0.96	1.60	2.27	1.00	1.07	1.39	1.81	0.56	1.38	
BAZ2A	J3QK86	NaN	NaN	0.54	0.77	0.34	1.11	NaN	NaN	0.72	0.48	0.85	0.58	0.50	NaN	NaN	NaN	0.34	1.83	NaN	NaN	0.60	0.94	NaN	NaN	0.83	NaN	
BCLAF1	E9PK91	0.23	0.24	NaN	NaN	NaN	NaN	0.81	NaN	NaN	NaN	NaN	NaN	NaN	2.21	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.41	NaN	NaN	
BCLAF1	Q9NYF8-2	9.47	0.39	3.98	3.10	0.69	0.83	0.80	1.37	0.85	0.66	4.28	2.20	NaN	NaN	0.98	0.78	0.89	0.97	1.13	1.31	0.58	0.72	1.38	1.51	0.68	0.79	
BCR/ABLf.	A9UEZ6	0.91	NaN	NaN	NaN	4.19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.11	NaN	NaN	NaN	
BCR/ABLf.	P11274-2	1.12	0.65	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.92	0.13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
BHLHB2	Q6IB83	0.62	0.59	0.56	0.36	0.46	0.55	0.46	0.36	0.45	0.50	0.43	0.47	NaN	NaN	0.68	1.19	0.39	0.45	1.78	0.26	0.40	0.51	0.32	1.34	0.68	0.30	
BLMH	Q13867	NaN	NaN	NaN	NaN	0.47	NaN	0.17	NaN	40.10	NaN	0.73	NaN	NaN	NaN	NaN	11.03	5.54	NaN	0.45	0.77	NaN	NaN	4.60	NaN	NaN	NaN	
BPTF	E9PE19	NaN	NaN	NaN	NaN	NaN	0.74	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.71	0.66	NaN	NaN	NaN	NaN	
BPTF	F5GXF5	NaN	NaN	NaN	0.06	0.56	0.32	NaN	NaN	0.34	0.20	0.31	0.32	NaN	NaN	NaN	NaN	0.32	0.27	NaN	NaN	0.25	0.33	NaN	NaN	NaN	0.58	
BPTF	F5H176	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.40	NaN	NaN	NaN	NaN	0.86	0.79	NaN	NaN	NaN	NaN	
BPTF	Q12830	0.29	1.33	0.54	0.79	0.64	0.67	0.57	0.58	0.93	0.45	0.72	0.63	0.37	1.10	0.61	0.75	0.70	0.64	2.01	0.39	0.63	0.61	1.04	0.71	0.74	0.44	
BPTF	Q12830-2	0.22	NaN	NaN	NaN	NaN	1.32	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.03	NaN	NaN	0.88	0.91	NaN	NaN	NaN	NaN	
BRCA1	Q3LRH8	0.86	0.39	0.31	4.37	0.25	2.39	0.78	0.65	1.09	0.62	0.95	0.85	NaN	1.03	0.43	0.12	0.75	0.91	1.18	1.10	0.23	2.40	1.11	0.47	0.62	0.72	
BRCC3	P46736-2	0.55	0.21	NaN	NaN	0.42	1.68	0.43	NaN	1.37	0.48	0.75	0.88	0.61	0.88	0.33	0.27	0.98	0.65	NaN	0.29	0.17	2.71	NaN	0.38	0.64	0.79	
BRD2	H0Y6K2	0.29	3.05	1.22	1.45	1.29	2.47	1.13	1.42	1.26	1.38	1.15	2.12	0.83	2.19	0.87	1.67	1.44	1.55	1.39	1.80	1.30	2.09	1.86	1.94	1.09	1.76	
BRE	Q9NXR7-4	0.48	0.26	0.21	3.04	0.42	1.66	0.61	NaN	1.46	0.45	0.75	0.70	0.61	NaN	0.41	0.12	0.97	0.63	0.63	0.47	0.15	3.33	0.53	0.25	0.59	0.62	
BRMS1	Q9Y3T1	0.59	0.81	1.51	1.22	0.71	2.90	1.53	1.53	1.35	1.64	1.29	1.06	0.46	NaN	NaN	1.89	1.07	1.44	10.22	0.23	1.04	1.55	0.64	4.30	NaN	NaN	
BRMS1L	B3KU43	0.80	1.03	1.51	1.56	0.81	3.14	1.74	1.25	1.68	1.92	1.66	1.33	0.63	3.27	0.51	0.87	1.30	1.77	6.12	0.36	1.22	2.08	0.82	4.80	3.32	0.97	
C10orf12	Q8N655	0.25	2.70	NaN	NaN	0.79	0.65	NaN	NaN	NaN	NaN	0.61	NaN	NaN	NaN	0.96	0.59	NaN	0.66	1.74	2.39	0.94	0.39	1.09	1.90	0.54	NaN	
C11orf57	Q6ZUT1	NaN	2.82	2.96	0.88	1.45	1.13	1.50	2.01	1.41	1.62	0.69	2.60	NaN	NaN	NaN	NaN	1.37	2.17	NaN	1.86	1.19	2.08	1.17	1.62	1.14	2.26	
C19orf147	Q8N9M1-2	NaN	NaN	NaN	NaN	0.46	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.86	6.53	0.33	0.51	0.24	NaN	NaN	NaN	NaN	
C9orf78	Q9NZ63	NaN	1.16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.18	NaN	NaN	NaN	NaN	0.37	NaN	NaN	NaN	NaN	NaN	
CACYBP	Q9HB71-3	1.44	NaN	NaN	NaN	1.13	0.59	0.65	NaN	1.12	0.71	0.61	0.66	0.68	0.86	1.42	0.65	0.52	NaN	NaN	NaN	0.67	0.68	NaN	0.86	NaN	NaN	
CAPN1	P07384	NaN	0.05	0.38	0.49	0.27	0.19	0.31	0.33	0.34	0.67	NaN	NaN	NaN	NaN	0.37	6.16	1.45	0.76	NaN	NaN	0.55	0.66	NaN	NaN	NaN	0.45	
CAPRIN1	Q14444-2	NaN	4.34	NaN	NaN	NaN	NaN	0.50	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.29	1.32	NaN	NaN	NaN	NaN	NaN	0.57	NaN	NaN	NaN	NaN	
CAST	B7Z468	NaN	0.11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CAST	E7EVY3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CAT	P04040	NaN	NaN	NaN	NaN	NaN	NaN	0.10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.08	NaN	NaN	NaN	NaN	NaN	NaN	1.01	
CBFB	Q13951	0.85	2.02	1.19	0.78	0.71	1.23	NaN	NaN	NaN	NaN	0.24	1.06	NaN	NaN	1.09	NaN	NaN	0.54	1.24	7.90	0.12	1.02	1.10	0.71	1.46	NaN	NaN
CBFB	Q13951-2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.39	0.94	NaN	NaN	
CBR1	P16152	NaN	0.28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.35	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.43	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
CBX1	Q6IBN6	NaN	NaN	1.17	0.74	0.77	1.22	0.78	0.79	0.68	1.04	1.04	0.89	1.18	0.62	1.84	0.42	7.84	0.07	26.41	0.16	8.65	0.06	NaN	0.80	2.66	0.32	
CBX3	A4D177	0.68	0.50	1.35	0.89	0.87	1.53	1.06	0.99	0.96	1.48	1.32	1.08	1.69	0.78	1.76	0.23	8.64	0.08	16.22	0.01	8.76	0.10	1.55	0.92	2.83	0.25	
CBX5	P45973	NaN	NaN	0.68	NaN	0.65	0.69	0.46	0.87	0.42	0.97	0.66	0.59	2.29	0.26	4.80	0.14	9.47	0.13	20.30	0.06	10.24	0.08	1.32	0.92	NaN	NaN	
CBX8	Q9HC52	0.77	NaN	1.23	0.97	1.10	1.26	0.97	1.00	2.61	0.51	3.58	0.39	0.82	1.92	0.53	0.21	1.89	0.61	6.76	0.30	2.31	0.56	2.05	1.16	1.86	0.91	
CCDC138	Q96M89	0.26	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.21	0.20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CCDC59	Q9P031	0.49	NaN	1.16	1.23	1.18	1.23	NaN	NaN	1.22	1.03	1.09	1.23	0.87	NaN	0.36	0.33	1.12	0.53	NaN	1.01	1.02	1.43	0.96	1.17	1.43	1.38	
CD3EAP	A8K818	NaN	NaN	NaN	0.79	0.61	0.58	0.75	0.59	NaN	NaN	1.42	0.61	NaN	NaN	NaN	6.25	0.32	0.95	5.46	0.44	1.62	0.28	NaN	NaN	NaN	NaN	
CD3EAP	O15446	1.47	1.84	0.73	0.90	0.62	0.65	1.16	0.79	0.33	2.02	1.49	0.58	NaN	NaN	1.22	7.41	0.30	0.94	2.77	0.35	1.54	0.30	NaN	NaN	NaN	NaN	
CDK2AP1	O14519-2	NaN	NaN	NaN	NaN	1.31	1.28	NaN	NaN	NaN	1.40	NaN	NaN	NaN	NaN	NaN	NaN	0.88	1.51	0.03	NaN	1.04	1.83	5.32	0.11	NaN	NaN	
CDKN1A	Q96LE1	NaN	NaN	NaN	0.53	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.40	NaN	NaN	NaN	0.59	NaN	NaN	NaN	NaN	
CDSN	Q0EFA5	0.08	NaN	NaN	NaN	NaN	NaN	0.14	NaN	0.21	0.06	0.25																

Gene names	Protein IDs	H3Δ1-20_F	H3Δ1-20_R	H4K20me1_F	H4K20me1_R	H4K20me3_F	H4K20me3_R	H3Kc27me1_F	H3Kc27me1_R	H3Kc27me2_F	H3Kc27me2_R	H3Kc27me3_F	H3Kc27me3_R	H3K9me1_F	H3K9me1_R	H3K9me2_F	H3K9me2_R	H3K9me3_F	H3K9me3_R	H3K9me3 meCpG_F	H3K9me3 meCpG_R	H3K9me20me3_F	H3K9me20me3_R	meCpG_F	meCpG_R	H4R3me2_F	H4R3me2_R	
CEBPD	P49716	NaN	0.23	1.60	0.91	0.60	2.75	NaN	NaN	NaN	1.92	1.26	1.32	NaN	NaN	0.10	1.23	0.48	2.77	5.85	0.15	1.69	0.92	1.26	1.37	NaN	NaN	
CEBPG	P53567	NaN	NaN	1.47	0.79	0.60	NaN	NaN	NaN	NaN	NaN	1.11	1.32	NaN	NaN	NaN	2.99	0.60	2.45	NaN	NaN	1.85	NaN	NaN	NaN	2.14	NaN	
CENPF	P49454	1.08	0.70	1.29	0.70	1.76	0.58	0.17	1.03	1.12	0.69	0.73	1.15	NaN	NaN	1.31	0.60	1.74	0.63	0.22	12.26	0.78	1.20	1.18	1.42	0.52	1.52	
CENPV	Q7Z7K6	1.87	1.97	0.87	0.65	0.88	0.92	0.84	0.80	0.77	0.86	0.78	0.77	0.72	0.97	2.21	4.94	0.73	0.75	1.40	1.22	0.89	0.77	0.92	1.02	0.77	1.14	
CENPV	Q7Z7K6-3	NaN	2.18	NaN	NaN	NaN	NaN	0.93	0.92	NaN	NaN	NaN	NaN	NaN	NaN	2.26	NaN	NaN	NaN	NaN	0.83	1.22	1.13	0.75	1.28	NaN	NaN	
CGGBP1	Q9UFW8	0.71	1.11	0.44	NaN	2.30	0.45	NaN	NaN	1.03	0.90	0.96	0.44	NaN	NaN	0.45	0.53	1.67	0.35	6.52	0.18	1.32	0.61	0.50	2.13	NaN	NaN	
CHAF1A	D6W625	0.49	1.07	0.96	0.48	1.34	0.83	0.55	0.82	0.41	1.55	0.76	0.92	NaN	0.64	0.86	1.02	4.77	0.12	5.08	0.37	6.15	0.10	1.50	1.54	NaN	NaN	
CHAF1B	B2R7X3	0.58	0.82	0.89	0.41	1.10	0.83	0.54	0.78	0.38	1.54	0.81	0.75	0.71	0.59	0.85	1.61	4.40	0.13	2.30	0.59	7.41	0.07	1.30	1.25	2.10	0.18	
CHAMP1	Q96JM3	1.07	0.71	1.09	0.85	1.06	1.04	0.61	NaN	0.77	1.66	1.22	1.35	1.21	NaN	0.63	0.14	6.98	0.18	11.02	0.12	7.92	0.14	4.49	0.54	1.06	0.34	
CHD1	O14646-2	1.40	0.47	0.81	0.54	1.32	0.48	0.67	1.82	1.08	0.57	NaN	0.99	NaN	NaN	1.33	0.46	1.09	0.45	0.31	6.01	0.54	1.03	0.78	0.95	0.23	1.99	
CHD1L	Q86WJ1	NaN	0.51	NaN	NaN	NaN	0.38	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.35	0.69	0.56	NaN	NaN	0.34	0.68	0.31	NaN	0.63	0.40	NaN	NaN	
CHD3	H7C0J3	NaN	NaN	NaN	NaN	NaN	1.15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.68	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CHD3	Q12873	0.53	NaN	0.60	0.49	0.59	0.41	0.48	0.46	0.69	0.49	0.68	0.41	0.44	1.34	0.42	0.16	0.38	0.46	4.21	0.28	0.61	0.41	4.32	0.15	0.76	0.22	
CHD4	F5GWX5	0.58	1.00	0.80	0.75	0.89	0.66	0.73	0.73	0.91	0.65	0.87	0.71	0.56	1.06	0.79	0.41	1.12	0.52	11.85	0.03	1.31	0.44	3.94	0.19	1.17	0.43	
CHD4	Q14839	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.60	0.53	8.52	0.26	2.15	0.32	NaN	NaN	NaN	NaN	NaN	
CHERP	Q8IWX8	1.76	2.12	0.86	1.82	1.60	0.83	0.84	1.16	1.52	1.37	NaN	1.66	NaN	NaN	1.26	0.62	1.61	0.94	1.38	1.03	0.88	2.02	1.16	1.13	0.91	1.36	
CHTOP	Q9Y3Y2-4	2.23	0.46	0.50	1.83	0.81	0.77	0.70	1.02	2.25	0.76	1.31	1.03	1.19	NaN	0.77	0.23	0.92	0.73	3.44	1.26	0.99	1.06	2.32	NaN	0.80	0.49	
CIR	A0PJ17	NaN	NaN	1.43	0.23	NaN	NaN	NaN	NaN	0.65	0.69	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.09	0.84	NaN	NaN	1.05	1.45	NaN	NaN	
CKAP2	B2RMQ4	3.76	2.14	0.41	0.64	0.53	0.50	0.55	NaN	0.64	0.71	0.72	0.92	0.53	NaN	0.01	0.25	0.53	0.75	0.35	1.36	0.50	0.71	0.87	0.67	0.38	0.52	
CKAP2	E9PD90	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.63	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CKB	P12277	NaN	0.18	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.30	NaN	NaN	NaN	0.29	NaN	NaN	NaN	1.18	0.39	NaN	NaN	
CLASP1	B7ZLX3	NaN	NaN	NaN	NaN	1.59	0.70	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.42	0.03	0.75	0.55	NaN	NaN	1.40	0.50	NaN	NaN	NaN	NaN	
CLASP1	Q7Z460	0.53	3.25	0.69	0.42	1.29	0.52	0.74	0.52	0.87	0.47	0.82	0.35	1.54	0.55	0.54	0.08	0.79	0.51	2.96	0.31	1.04	0.40	0.53	0.67	1.16	0.25	
CLASP2	F5H604	0.49	2.66	0.70	0.54	1.54	0.54	0.91	0.68	0.66	0.73	1.01	0.45	1.73	0.60	0.51	0.08	1.04	0.58	1.92	0.25	0.70	0.77	0.69	NaN	1.24	0.45	
CLIC1	Q5SRT3	0.18	0.15	NaN	NaN	NaN	NaN	NaN	NaN	1.79	0.32	NaN	NaN	NaN	NaN	0.25	NaN	NaN	NaN	3.18	1.79	NaN	NaN	1.15	0.31	NaN	NaN	
CLOCK	Q53EU0	1.24	NaN	0.91	NaN	1.62	2.15	1.18	1.14	0.81	1.26	1.25	NaN	NaN	2.57	NaN	NaN	1.40	1.38	0.10	0.40	1.52	1.38	0.63	2.19	1.67	1.54	
CLTC	K7EJ5	NaN	NaN	NaN	1.51	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CLUH	I3L2B0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CLUH	K7EI61	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CNBP	P62633-2	1.04	0.79	1.07	0.57	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.78	2.12	NaN	NaN	0.88	0.73	0.50	NaN	1.24	NaN	NaN	NaN	
CNOT1	A5YKK6	0.91	1.60	0.56	0.79	1.01	0.30	1.05	0.34	0.52	0.68	0.46	0.57	1.42	0.33	0.76	0.06	1.17	0.67	2.19	NaN	NaN	0.48	0.96	NaN	NaN	1.10	0.47
CNOT1	H3BMZ2	1.05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CNOT10	Q9H9A5	0.87	1.17	0.48	0.66	1.07	0.30	0.86	NaN	0.36	0.64	0.42	0.55	1.16	0.48	0.72	0.18	1.05	0.69	0.82	NaN	0.50	0.84	NaN	NaN	0.86	0.51	
CNOT11	Q9UKZ1	0.67	0.77	0.71	0.61	0.88	0.40	1.23	0.42	0.52	0.77	0.51	0.46	1.02	0.42	0.63	0.15	1.20	0.72	1.30	NaN	0.56	0.95	NaN	NaN	1.33	0.55	
CNOT2	B3KTL6	NaN	NaN	0.81	0.91	0.85	0.83	NaN	0.95	0.92	0.98	0.82	0.85	0.89	1.33	NaN	NaN	0.75	0.96	NaN	NaN	0.89	0.78	0.85	0.78	0.72	0.72	
CNOT2	F8VV52	0.48	0.97	0.48	0.55	0.87	0.32	1.27	0.32	0.59	0.86	NaN	NaN	0.79	0.55	0.53	0.21	0.93	0.61	NaN	NaN	0.37	0.86	NaN	NaN	0.84	0.52	
CNOT3	H7C148	NaN	NaN	NaN	NaN	NaN	NaN	0.63	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CNOT7	B3KM57	0.86	1.18	0.73	0.75	1.03	0.52	1.19	0.45	0.60	1.02	0.53	0.66	0.94	0.49	0.69	0.06	1.08	0.72	1.87	NaN	0.62	1.08	NaN	NaN	1.72	0.57	
COBL	O75128-3	0.37	1.68	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.70	NaN	NaN	0.32	0.05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.39	
CPNE1	F2Z2V0	NaN	0.30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.65	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CPNE8	Q86YQ8	NaN	2.88	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.08	NaN	NaN	0.62	1.26	NaN	NaN	NaN	0.94	NaN	NaN	
CRCP	O75575	NaN	NaN	NaN	NaN	0.44	0.62	0.35	0.75	0.51	0.75	0.67	0.66	NaN	NaN	0.34	NaN	0.20	1.42	11.02	NaN	1.44	0.25	NaN	NaN	NaN	NaN	
CSTA	Q6IB90	NaN	NaN	0.55	NaN	NaN	NaN	0.07	NaN	0.40	NaN	NaN	0.57	NaN	NaN	0.08	NaN	NaN	NaN	0.06	NaN	0.08	NaN	NaN	0.21	NaN	NaN	
CTSD	P07339	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.18	NaN	NaN	NaN	NaN	0.18	NaN	NaN	
CUEDC1	J3QLQ8	NaN	NaN	NaN	NaN	1.84	0.61	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.52	NaN	NaN	NaN	0.76	1.05	13.24	0.16	NaN	NaN	
CUL1	Q13616	4.92	0.60	1.49	NaN	1.57	1.17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.74	NaN	NaN	NaN	0.99	1.93	0.62	1.24	NaN	1.07	0.92	NaN	NaN	
CXorf38	B2RD30	4.76	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
CYR61	B4DI61	12.33	9.81	0.73	0.75	0.61	0.47	0.54	NaN	NaN	NaN	0.67	0.94	0.36	NaN	9.50	7.97	0.33	NaN	1.11	1.14	NaN	0.55	1.04	0.96	NaN	NaN	
DDIT4	Q9NX09	0.35	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.39	0.12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
DDX21	Q9NR30	1.52	3.92	0.33	1.51	1.03	1.03	1.05	NaN	1.51	0.76	1.36	0.66	0.85	1.17	2.92	0.83	0.72	0.81	1.07	1.67	0.75	0.91	2.31	1.25	0.79	1.11	
DDX52	Q9Y2R4	0.44	2.36	1.35	1.06	1.11	0.83	0.91	NaN	1.63	0.95	0.95	1.25	1.62	0.77	0.54	0.15	1.00	0.92	1.44	0.94	1.18	1.00	1.65	0.93	0.90	2.38	
DDX5-ETV4 f.	C1IK54	0.45	NaN	NaN	NaN	0.11	0.31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.17	NaN	NaN	0.04	0.28	NaN	NaN	NaN	NaN	NaN	NaN	
DEFA3	Q6EZ9	0.02	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
DFNA5	H7C147	0.17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
DFNA5	G60443	0.41	0.15	0.66	0.81	0.73	0.25	1.17	0.35	0.54	0.71	0.38	0.52	0.68	0.88	0.34	0.13	1.45	0.5									

Gene names	Protein IDs	H3Δ1-20_F	H3Δ1-20_R	H4K20me1_F	H4K20me1_R	H4K20me3_F	H4K20me3_R	H3Kc27me1_F	H3Kc27me1_R	H3Kc27me2_F	H3Kc27me2_R	H3Kc27me3_F	H3Kc27me3_R	H3K9me1_F	H3K9me1_R	H3K9me2_F	H3K9me2_R	H3K9me3_F	H3K9me3_R	H3K9me3 meCpG_F	H3K9me3 meCpG_R	H3K9me3 H4K20me3_F	H3K9me3 H4K20me3_R	meCpG_F	meCpG_R	H4R3me2_F	H4R3me2_R
DNMT1	F5GX68	0.19	1.69	1.18	0.42	2.58	0.46	0.60	0.96	0.83	0.78	1.04	0.69	1.45	0.46	4.40	0.28	3.14	0.18	56.86	0.01	1.82	0.38	31.78	0.02	0.38	1.61
DNMT1	P26358	NaN	NaN	NaN	1.48	0.38	2.75	0.50	NaN	NaN	NaN	0.95	0.70	1.41	NaN	NaN	NaN	3.32	0.20	NaN	NaN	1.96	0.44	NaN	NaN	NaN	NaN
DNMT1	P26358-2	NaN	NaN	1.29	0.36	2.44	0.44	NaN	NaN	NaN	NaN	1.05	0.66	NaN	0.23	NaN	NaN	2.92	0.18	215.33	0.01	1.88	0.36	39.19	0.05	NaN	NaN
DSC1	Q9HB00	0.08	0.07	NaN	NaN	0.13	NaN	0.06	NaN	0.37	0.09	NaN	NaN	0.68	1.08	0.08	NaN	NaN	NaN	0.06	0.13	0.13	NaN	0.17	NaN	NaN	NaN
DSC2	Q68DY8	NaN	NaN	NaN	NaN	NaN	NaN	0.21	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
DSC3	A8K6T3	NaN	NaN	NaN	NaN	NaN	1.00	0.12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.11	NaN	NaN	NaN	NaN	NaN	NaN	NaN
DYM	J3QSE7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
DYM	Q7RTS9	0.79	0.91	0.44	0.72	0.91	0.32	1.11	0.34	0.44	0.65	0.45	0.38	0.76	0.51	0.73	0.14	1.28	0.71	0.71	NaN	0.49	0.99	NaN	NaN	1.27	0.51
E2F1	Q01094	NaN	NaN	NaN	NaN	0.65	0.50	0.62	NaN	0.46	0.46	0.54	0.74	NaN	NaN	NaN	NaN	0.52	0.36	NaN	0.38	0.51	0.45	0.23	1.85	NaN	NaN
E2F2	Q5U0J0	NaN	NaN	NaN	NaN	1.04	0.74	NaN	NaN	0.58	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.65	0.51	NaN	NaN	0.57	0.65	0.45	1.49	NaN	NaN
E2F6	Q75461	0.75	NaN	3.04	1.67	1.61	3.48	2.27	2.24	1.77	3.06	2.02	2.42	1.10	2.77	0.78	0.49	2.47	1.83	2.05	0.13	3.44	1.48	1.71	3.74	5.28	NaN
E4F1	H3BUJ7	NaN	0.51	1.28	0.77	0.34	3.30	0.65	NaN	0.58	2.27	1.28	1.30	0.27	6.34	0.08	4.58	0.21	4.35	12.56	0.10	1.64	0.73	0.93	2.05	NaN	NaN
EED	Q75530	0.29	1.91	0.89	0.64	0.90	0.66	0.65	0.90	0.81	0.74	0.76	0.77	0.50	1.51	1.31	0.76	1.46	0.33	2.34	0.29	1.41	0.35	1.01	1.60	0.71	0.60
EEF1B2	P24534	0.65	0.34	1.08	NaN	0.52	1.02	0.40	2.66	0.76	0.49	0.79	0.61	NaN	NaN	NaN	0.80	0.33	1.20	0.66	0.87	0.90	0.57	1.35	NaN	1.22	NaN
EEF2	P13639	0.88	0.41	1.39	NaN	1.10	0.97	0.40	NaN	1.35	0.60	1.12	NaN	1.50	0.93	0.74	0.39	1.02	0.85	0.18	1.50	0.61	0.86	0.97	0.38	NaN	0.88
EHMT1	Q9H9B1	0.45	0.51	1.05	0.92	0.68	1.60	0.85	1.06	0.94	1.27	1.14	0.98	2.16	0.74	1.03	0.25	0.38	2.02	8.61	0.20	1.05	0.87	2.20	2.01	2.29	0.34
EHMT2	A2ABF8	0.54	0.56	0.92	0.93	0.72	1.64	0.82	0.95	0.92	1.15	1.11	0.92	2.24	0.73	1.24	0.48	0.37	1.94	7.75	0.23	1.08	0.85	2.24	1.91	1.68	0.28
EIF2B5	Q13144	NaN	0.73	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	22.40	0.95	1.34	1.12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
EIF2S1	Q53XC0	1.52	2.07	1.09	1.30	8.22	0.34	0.89	1.13	1.29	0.76	1.96	0.50	0.55	0.46	1.32	1.23	2.84	0.35	NaN	0.26	0.97	1.00	1.09	2.28	1.57	0.64
EIF2S2	Q6IBR8	0.93	1.39	0.71	0.83	8.30	0.33	1.11	NaN	1.28	0.74	2.18	0.46	0.45	0.43	1.18	2.24	2.71	0.33	2.23	0.28	0.95	1.08	0.90	1.06	1.50	0.53
EIF2S3	P41091	0.82	1.35	1.01	1.02	7.23	0.32	0.88	1.04	1.16	0.67	2.08	0.52	0.79	0.72	1.15	0.98	2.37	0.34	2.61	0.62	0.88	1.03	1.23	1.24	0.95	0.52
EIF4A3	P38919	4.31	0.92	0.48	1.40	0.83	0.61	0.79	0.90	2.11	0.30	1.08	0.97	1.37	0.99	2.02	1.56	1.05	0.56	1.37	2.38	0.58	1.24	1.96	1.34	0.56	0.55
EIF4G1	E7EX73	1.30	1.24	NaN	NaN	0.66	0.62	NaN	0.71	1.14	0.38	NaN	NaN	NaN	0.61	1.50	0.14	0.63	0.74	1.17	1.34	0.35	NaN	NaN	0.88	0.62	NaN
EIF4H	Q15056	NaN	0.20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.36	0.56	NaN	NaN	0.43	1.00	NaN	NaN	0.89	0.87	NaN	NaN
EIF5	P55010	NaN	0.40	NaN	NaN	15.49	0.26	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.76	0.19	NaN	0.28	NaN	2.89	0.73	0.68	NaN	NaN
EIF5B	Q8N5A0	NaN	NaN	NaN	NaN	6.17	0.19	NaN	0.86	NaN	NaN	0.86	0.62	NaN	NaN	NaN	NaN	0.88	0.21	NaN	0.24	0.39	1.02	NaN	1.53	NaN	NaN
ELF1	P32519-2	0.80	NaN	NaN	1.02	3.68	0.40	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.19	3.47	2.27	0.34	3.10	0.24	2.38	0.57	0.47	1.85	NaN	NaN
ELF2	B7Z720	0.69	0.40	NaN	NaN	3.04	0.66	NaN	NaN	NaN	NaN	1.99	0.55	NaN	NaN	0.12	0.73	1.85	0.64	8.91	0.17	2.59	0.55	0.58	2.58	NaN	NaN
ELF2	Q15723-1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
EMG1	Q92979	0.51	0.79	0.84	0.87	7.42	0.27	1.02	0.93	1.61	2.89	1.04	0.62	0.75	0.62	0.53	0.15	2.79	0.25	2.15	0.49	0.72	0.96	1.43	1.85	0.85	NaN
ENO1	P06733	0.18	0.07	3.44	NaN	1.91	0.98	0.25	2.61	1.90	0.36	NaN	NaN	NaN	1.03	0.16	0.16	0.79	NaN	0.32	1.67	0.46	0.45	1.01	0.27	NaN	NaN
EPPK1	P58107	NaN	0.69	1.42	1.10	1.01	0.48	0.13	NaN	NaN	0.38	0.70	1.68	0.75	0.41	0.75	0.57	0.63	0.57	NaN	1.30	0.45	0.58	0.16	NaN	NaN	NaN
ESRRA	Q569H8	NaN	NaN	NaN	0.56	2.28	0.83	0.83	0.95	0.25	3.92	0.78	0.87	NaN	NaN	NaN	NaN	0.80	1.42	4.48	0.22	1.82	0.54	NaN	NaN	1.46	0.50
ETV6	P41212	0.57	0.56	1.66	1.44	0.83	2.60	1.27	1.96	0.97	2.56	1.31	1.17	0.78	2.38	0.27	0.71	0.65	2.49	10.92	0.03	1.86	0.84	0.80	1.04	2.01	1.12
EXOC2	Q96KP1	0.88	1.19	0.84	0.95	0.94	0.53	1.13	0.70	0.94	0.99	0.74	1.46	0.77	0.64	0.77	0.21	1.33	1.05	NaN	NaN	0.65	1.22	NaN	NaN	2.44	0.60
EZH2	Q15910	0.32	1.99	0.77	0.71	1.06	0.94	0.73	0.94	0.87	0.98	0.90	1.03	0.59	1.42	1.34	1.02	1.49	0.45	1.85	0.94	1.38	0.42	1.13	1.69	0.99	0.74
FABP5	E7DWW5	NaN	NaN	NaN	NaN	NaN	NaN	0.07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.08	0.18	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
FAM114A1	Q8IWE2	NaN	0.05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
FAM129B	Q96TA1-2	1.31	1.04	NaN	NaN	NaN	0.85	0.50	0.80	NaN	NaN	NaN	1.24	NaN	NaN	1.46	0.93	0.29	NaN	0.26	NaN	0.90	1.12	0.92	1.09	0.01	0.56
FAM134C	Q86VR2	NaN	0.22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.62	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
FAM48A	B3KNI1	NaN	0.22	0.54	0.41	0.45	0.91	NaN	NaN	NaN	NaN	0.94	0.65	NaN	NaN	0.33	NaN	0.50	0.93	NaN	NaN	1.16	0.32	NaN	NaN	NaN	NaN
FBXL6	Q8N531	0.68	0.30	0.95	0.54	0.69	0.80	1.09	0.40	0.69	0.64	0.90	0.41	0.42	1.61	0.62	NaN	0.67	0.65	2.90	0.14	0.75	0.67	0.60	1.03	0.49	0.72
FBXW11	B4DH70	NaN	2.12	1.02	0.79	0.61	1.12	0.82	0.81	0.71	1.07	0.80	0.85	0.45	2.58	NaN	NaN	0.51	1.11	8.04	0.21	0.91	0.77	1.16	1.11	NaN	0.53
FGF2	D9ZGF5	0.88	1.98	0.50	1.65	0.99	0.84	0.81	NaN	1.21	4.17	4.84	NaN	NaN	NaN	1.59	4.47	0.82	0.72	0.79	0.41	0.70	0.81	0.86	0.46	0.44	0.76
FIZ1	Q96SL8	0.81	0.39	0.64	0.82	0.46	1.52	0.80	0.77	0.52	0.64	0.71	0.62	0.29	1.40	0.62	6.13	0.28	1.13	9.62	0.04	0.58	0.50	6.84	0.17	1.19	0.30
FMNL1	I6L919	NaN	0.15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.30	0.63
FMNL1	Q95466-2	0.86	0.13	0.70	1.00	0.87	0.34	NaN	NaN	NaN	NaN	NaN	0.84	NaN	NaN	0.62	2.51	1.12	1.00	1.21	0.99	1.05	1.18	NaN	NaN	NaN	0.60
FNBP4	D3DQS4	NaN	2.29	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.98	0.96	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
FOXN2	Q6IS90	0.42	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.84	1.17	0.78	NaN	0.34	NaN	NaN	NaN	NaN	NaN	0.70	NaN	NaN	NaN	NaN	NaN
FRMPD3	Q5JV73	0.07	NaN	0.95	0.85	0.29	0.88	NaN	NaN	NaN	0.04	7.21	0.78	7.64	1.26	0.94	NaN	0.56	0.21	0.38	0.09	0.40	2.45	0.10	NaN	0.33	0.25
G2E3	F5GX24	0.34	0.38	1.30	1.13	1.23	0.81	0.25	2.63	0.56	1.38	0.29	4.19	1.36	1.05	0.76	0.67	1.22	0.79	0.46	12.76	0.94	1.18	2.66	1.57	0.66	1.18
GATAD2A	Q86YP4	0.39	0.90	1.01	0.72	0.95	0.84	0.98	0.96	1.01	0.82	0.92	0.85	0.41	1.45	0.67	0.87	0.68	1.10	5.61	0.05	0.69	1.03	4.53	0.15	1.45	0.52
GATAD2B	Q8WX99	0.43	0.85	0.81	0.61	0.85	0.61	0.66	0.72	0.82	0.69	0.76	0.68	0.34	1.32	0.69	0.50	0.58	0.92	5							

Gene names	Protein IDs	H3Δ1- 20_F	H3Δ1- 20_R	H4K20 me1_F	H4K20 me1_R	H4K20 me3_F	H4K20 me3_R	H3Kc27 me1_F	H3Kc27 me1_R	H3Kc27 me2_F	H3Kc27 me2_R	H3Kc27 me3_F	H3Kc27 me3_R	H3K9m e1_F	H3K9m e1_R	H3K9m e2_F	H3K9m e2_R	H3K9m e3_F	H3K9m e3_R	H3K9m e3 meC pG_F	H3K9m e3 meC pG_R	H3K9m 20me3_F	H3K9m e3 H4K 20me3_R	meCpG _F	meCpG _R	H4R3m e2_F	H4R3m e2_R	
GGNBP2	A8K3S2	0.44	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.63	0.46	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GNB2L1	E9KL35	1.08	1.05	NaN	NaN	1.24	0.99	0.25	NaN	NaN	0.43	NaN	NaN	0.64	1.33	1.01	0.23	1.26	NaN	NaN	1.02	1.28	1.31	0.68	0.68	1.40	NaN	
GOLGB1	Q14789	NaN	0.24	0.08	0.42	0.60	0.12	NaN	NaN	NaN	NaN	0.05	NaN	NaN	1.33	NaN	NaN	1.12	0.67	NaN	NaN	0.67	0.81	NaN	NaN	0.97	1.29	
GOLTI1B	G3V1U5	NaN	NaN	NaN	NaN	NaN	0.19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.04	NaN	1.39	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
GSTP1	P09211	NaN	NaN	NaN	NaN	NaN	NaN	0.10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.70	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
GTF2E2	P29084	NaN	NaN	NaN	0.59	1.47	0.50	0.79	2.24	0.94	1.66	0.57	2.17	0.31	3.64	0.68	0.79	NaN	NaN	NaN	NaN	1.66	0.95	NaN	1.73	NaN	NaN	
GTF3C4	Q9UKN8	0.81	0.55	0.51	0.55	0.71	0.66	0.51	NaN	0.51	0.89	0.81	0.56	0.53	0.95	0.80	0.33	0.83	0.46	5.58	0.34	1.82	0.17	0.80	NaN	NaN	NaN	
GTF3C5	Q5T7U4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
GTF3C5	Q9Y5Q8	0.62	0.47	0.45	0.72	0.66	0.53	0.47	NaN	0.37	0.84	0.70	0.51	0.40	1.24	0.68	0.18	0.96	0.37	9.10	0.11	2.15	0.17	0.81	1.55	NaN	0.69	
HDGF	P51858	NaN	0.10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
HDGFRP2	Q7Z4V5-2	NaN	0.57	8.74	0.20	0.70	0.77	NaN	0.65	0.78	NaN	0.52	NaN	NaN	NaN	NaN	NaN	0.92	0.47	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
HEATR2	Q86Y56	1.08	0.72	0.88	0.99	1.21	0.53	1.72	NaN	0.20	1.30	0.42	0.85	1.14	0.98	0.93	0.56	1.86	0.99	0.33	NaN	0.67	1.21	NaN	NaN	2.21	0.77	
HES1	Q14469	0.79	1.39	2.08	0.44	1.02	1.41	1.23	1.26	0.82	1.50	1.13	1.04	NaN	NaN	0.52	0.95	0.79	1.38	NaN	0.22	1.53	0.81	0.44	4.35	3.09	NaN	
HIVEP2	P31629	1.30	3.32	0.69	NaN	0.44	1.34	NaN	NaN	NaN	0.63	0.52	0.55	NaN	NaN	NaN	NaN	0.42	NaN	4.31	0.58	0.73	0.42	NaN	NaN	NaN	NaN	
HMG20A	Q9NP66	0.12	1.27	1.51	1.00	1.16	1.34	1.07	1.23	1.07	1.59	1.13	1.29	NaN	NaN	0.49	0.64	1.09	1.16	0.84	1.42	1.11	1.02	1.01	1.12	1.17	1.90	
HMG41	P17096	2.15	2.37	0.55	0.93	0.93	0.78	0.54	0.54	0.67	0.59	0.91	0.65	0.69	0.76	2.09	2.24	0.92	0.64	1.28	1.31	0.75	0.64	1.37	1.27	0.49	0.90	
HMG83	O15347	0.30	1.62	1.14	1.80	5.09	0.89	1.06	1.57	1.47	0.95	0.46	NaN	NaN	NaN	0.28	NaN	1.23	0.50	4.72	0.91	1.25	0.81	2.74	1.92	NaN	0.93	
HMGN5	Q5JSL0	NaN	NaN	NaN	NaN	18.80	0.09	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.10	0.19	NaN	NaN	NaN	NaN	NaN	NaN	0.70	NaN	
HS90AB1	P08238	1.29	0.95	1.17	1.09	1.25	0.75	0.59	1.67	1.12	0.52	0.70	0.80	0.94	1.35	1.60	1.27	0.96	0.84	0.28	2.48	0.74	0.75	1.11	0.62	1.08	0.72	
HSPA9	B7Z4V2	1.19	1.28	0.48	0.59	0.50	0.51	0.48	NaN	0.40	0.55	0.39	0.43	NaN	0.81	1.11	0.97	0.39	0.50	NaN	0.48	0.57	0.31	NaN	NaN	NaN	2.30	
HTRA1	Q05DJ8	0.15	0.24	0.33	0.44	0.34	0.40	0.32	0.61	0.73	0.60	0.32	1.08	0.42	0.34	NaN	NaN	0.45	0.83	0.24	NaN	0.39	0.37	NaN	NaN	0.29	0.30	
HYDIN	Q4G0P3	NaN	NaN	NaN	1.73	0.79	2.22	1.34	1.54	NaN	NaN	1.72	1.47	NaN	NaN	NaN	NaN	1.14	1.69	0.78	0.18	1.58	1.39	4.78	0.12	1.03	NaN	
IFI16	D3DUZ3	0.26	1.71	0.45	1.04	0.58	0.89	0.44	0.63	0.69	0.69	0.85	0.72	0.49	1.09	0.32	7.31	0.59	0.78	3.99	0.04	0.66	0.68	0.96	0.44	0.87	0.30	
IFI16	Q16666	0.29	1.81	0.54	1.12	0.65	0.97	0.55	0.75	0.79	0.79	0.94	0.81	0.57	1.29	0.33	4.45	0.63	0.90	10.65	0.01	0.76	0.78	0.99	0.50	1.13	0.34	
IGF1	Q9NP10	0.06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.04	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
IGF2	E3JUN46	0.19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.34	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
IGHG1	S6B291	NaN	NaN	0.99	NaN	0.97	NaN	0.07	NaN	NaN	NaN	NaN	0.90	NaN	NaN	NaN	0.16	0.71	NaN	NaN	NaN	0.97	0.74	0.24	NaN	NaN	NaN	
IGHG4	P01861	NaN	NaN	NaN	NaN	NaN	NaN	0.29	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
IGKC	Q0KKI6	0.06	NaN	NaN	NaN	NaN	NaN	0.07	NaN	NaN	0.42	NaN	NaN	NaN	NaN	NaN	0.16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.28	NaN	0.57	
IGLC1	A2NUT2	NaN	NaN	NaN	NaN	NaN	NaN	0.04	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
ING1	Q9UK53-2	NaN	1.18	1.37	1.36	0.73	2.89	1.63	1.62	1.39	1.82	1.35	1.12	0.55	2.88	0.47	2.62	1.23	1.39	NaN	0.30	1.05	1.62	0.51	5.21	1.67	NaN	
ING2	B2RA15	NaN	0.71	1.53	1.50	0.94	3.35	1.92	2.15	1.69	2.05	1.62	1.44	0.69	3.35	NaN	NaN	1.59	1.92	NaN	0.25	1.60	2.09	0.66	6.47	2.70	1.47	
ING4	A4KYM7	NaN	NaN	0.76	0.81	0.68	0.93	0.73	0.90	1.16	0.88	0.69	1.07	0.52	1.94	0.65	NaN	0.98	0.77	0.36	NaN	0.25	2.33	1.43	2.34	1.02	1.22	
ING5	E9PEN0	NaN	NaN	1.14	1.15	1.32	1.80	1.10	1.06	1.48	1.03	NaN	NaN	0.97	NaN	0.70	NaN	1.81	1.58	0.52	6.29	0.61	3.50	1.95	2.74	1.55	1.48	
INO80	Q9ULG1	0.66	0.47	1.56	0.64	0.85	1.46	1.06	0.81	0.65	2.13	1.25	1.07	0.32	3.48	0.33	2.89	0.46	2.07	45.99	0.19	2.55	0.48	1.58	2.30	2.54	1.03	
INO80B	Q9C086	0.38	0.29	2.49	0.68	0.79	1.97	1.44	0.83	0.73	2.47	1.33	1.12	0.22	3.93	0.26	1.26	0.51	2.56	25.25	0.09	2.89	0.56	1.13	0.80	2.10	0.38	
INO80C	Q6PI98	0.71	0.40	2.94	0.48	0.80	2.02	1.68	0.97	0.72	3.02	1.50	1.27	0.26	4.42	0.25	1.33	0.55	2.66	18.02	0.07	3.32	0.59	1.08	1.64	2.83	0.47	
INO80D	Q53TQ3-2	NaN	NaN	2.54	0.56	0.82	1.15	NaN	NaN	0.72	2.81	1.48	1.42	NaN	NaN	NaN	NaN	0.56	2.04	NaN	0.25	3.22	0.53	NaN	NaN	NaN	NaN	
INO80E	Q8NBZ0	0.41	0.29	2.36	0.62	0.77	1.85	1.45	0.92	0.64	2.89	1.37	1.26	0.21	4.84	0.25	0.93	0.52	2.59	11.44	0.04	3.09	0.54	1.08	1.59	NaN	0.52	
IP011	Q9UI26	1.27	1.12	0.77	0.95	0.80	0.38	1.15	0.39	0.61	0.77	0.74	1.27	0.91	0.67	1.08	0.14	1.51	0.77	NaN	NaN	0.60	1.22	NaN	NaN	1.95	0.57	
ISL2	Q96A47	NaN	NaN	1.47	0.79	1.00	NaN	1.10	NaN	0.56	2.13	0.88	0.96	NaN	NaN	NaN	NaN	0.94	NaN	15.38	0.15	1.88	0.56	0.96	1.32	NaN	NaN	
JADE2	G3XAA4	0.19	0.37	0.99	1.05	0.87	1.15	8.73	1.31	14.63	0.96	0.77	1.85	NaN	1.35	NaN	NaN	0.72	1.12	NaN	NaN	0.46	2.27	NaN	NaN	NaN	1.37	
JADE3	Q92613	0.18	2.96	0.59	0.90	0.87	0.90	0.47	1.11	1.04	0.73	1.03	1.50	0.67	1.38	1.44	0.61	0.87	0.69	0.27	4.05	0.21	2.48	1.26	1.80	0.60	0.77	
JARID2	Q92833	NaN	NaN	0.42	0.96	0.68	0.90	NaN	NaN	0.50	0.87	0.18	2.49	NaN	1.09	NaN	NaN	0.56	0.83	NaN	0.58	0.64	0.66	1.42	1.15	NaN	NaN	
JMJDC1	B7ZLC8	1.20	NaN	NaN	NaN	1.26	0.41	NaN	NaN	NaN	0.78	0.74	NaN	0.65	NaN	0.53	0.06	1.17	0.25	0.02	NaN	0.77	0.40	NaN	NaN	NaN	NaN	
KAT7	Q95251-4	0.20	1.58	0.43	0.79	0.69	0.61	0.42	0.66	0.78	0.51	0.40	0.72	0.58	0.68	1.09	1.05	0.70	0.64	0.11	4.40	0.26	1.14	1.12	1.12	0.50	0.76	
KIAA1524	Q8TCG1	0.97	0.70	0.54	0.76	0.99	0.49	NaN	NaN	11.50	0.83	NaN	NaN	0.80	1.02	0.76	0.35	1.46	0.75	0.18	NaN	0.59	1.51	NaN	NaN	2.77	0.64	
KIF11	P52732	1.28	0.94	0.77	1.45	0.67	0.19	1.07	0.38	0.55	0.73	0.40	NaN	0.93	0.75	1.13	0.15	1.31	0.73	NaN	NaN	0.65	1.23	NaN	NaN	1.11	0.43	
KIF2A	O00139-1	NaN	NaN	NaN	NaN	NaN	0.40	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
KIF2A	O00139-2	NaN	NaN	0.69	NaN	5.41	0.26	NaN	NaN	NaN	NaN	1.25	0.40	NaN	NaN	0.66	0.41	1.97	0.25	1.53	0.58	0.94	0.68	1.14	1.05	NaN	0.58	
KIF2C	A8K3S3	1.01	0.78	NaN	0.69	4.69	0.23	NaN	NaN	NaN	0.45	0.79	NaN	NaN	0.89	NaN	0.33	1.03	0.48	NaN	0.90	0.61	0.65	NaN	NaN	NaN	0.45	
KIF5B-ALK	C1PHA2	NaN	NaN	1.16	NaN	NaN	NaN	NaN	0.59	0.28	NaN	0.73	NaN	NaN	NaN	NaN	1.56	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
KIFC1	Q9BWW19	0.62	0.57	0.37	NaN	2.36	0.20	NaN	NaN	0.61	0.25	0.73	0.28	NaN	0.43	0.42	0.12	1.25	0.19	1.86	NaN	0.6						

Gene names	Protein IDs	H3Δ1-20_F	H3Δ1-20_R	H4K20 me1_F	H4K20 me1_R	H4K20 me3_F	H4K20 me3_R	H3K9c27 me1_F	H3K9c27 me1_R	H3K9c27 me2_F	H3K9c27 me2_R	H3K9c27 me3_F	H3K9c27 me3_R	H3K9m e1_F	H3K9m e1_R	H3K9m e2_F	H3K9m e2_R	H3K9m e3_F	H3K9m e3_R	H3K9m e3/meC pG_F	H3K9m e3/meC pG_R	H3K9m e3/H4K20me3_F	H3K9m e3/H4K20me3_R	meCpG_F	meCpG_R	H4R3m e2_F	H4R3m e2_R
LDHA	P00338	0.34	0.17	1.63	1.45	1.42	1.11	0.45	2.65	1.51	0.58	1.13	1.24	1.24	1.02	0.27	0.22	1.24	1.08	0.78	1.94	0.82	1.00	1.41	0.72	2.07	1.10
LDHB	Q5U077	0.99	0.52	1.28	1.04	0.86	0.69	0.42	2.03	1.20	0.52	0.96	1.07	1.15	0.95	0.82	0.22	1.00	0.94	0.51	1.48	0.66	0.84	1.18	0.69	1.77	0.55
LGALS7	P47929	NaN	NaN	NaN	NaN	NaN	NaN	0.08	NaN	NaN	0.20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.06	NaN	NaN	NaN	NaN
LIG4	P49917	NaN	1.95	1.64	1.27	0.87	1.65	1.30	1.22	0.82	2.19	1.37	1.52	0.65	2.57	NaN	NaN	0.82	1.62	4.67	0.16	1.65	1.02	1.06	0.83	2.15	0.78
LMNA	P02545	0.69	0.42	1.15	0.57	1.99	0.35	0.72	0.49	1.10	0.86	0.78	1.06	0.96	1.12	0.56	0.62	1.24	0.60	0.07	2.21	0.79	0.97	0.10	0.56	0.19	1.73
LMNA	P02545-2	NaN	0.25	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.31	NaN	NaN	NaN	0.09	1.46	NaN	NaN	NaN	NaN	NaN	NaN
LMNA	P02545-5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LRIF1	Q5TJ3J	NaN	NaN	1.54	NaN	1.22	1.22	NaN	NaN	0.91	NaN	2.35	1.15	NaN	NaN	NaN	NaN	3.65	0.67	5.72	0.13	2.75	0.38	2.52	NaN	NaN	NaN
LRRFIP2	A8MXR0	NaN	4.13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.36	1.04	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LRWD1	Q9UFC0	0.76	0.71	1.12	1.09	9.82	0.05	1.47	1.05	0.91	1.15	3.02	0.19	0.99	0.93	0.78	0.25	3.16	0.18	3.42	0.61	7.94	0.05	2.90	1.09	1.35	1.11
LSM2	Q9Y333	1.43	0.87	NaN	NaN	1.31	0.73	NaN	1.09	NaN	NaN	NaN	NaN	NaN	NaN	0.95	0.40	NaN	0.77	0.61	1.13	0.07	1.50	NaN	NaN	0.63	0.70
MAFK	A2VCQ5	NaN	NaN	NaN	NaN	NaN	0.28	0.09	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MAFK	A8WFP5	0.87	NaN	0.70	0.67	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.51	NaN	NaN	0.34	1.88	NaN	NaN	5.77	0.12	NaN	0.57	2.28	0.49	NaN	NaN
MAGOH	P61326	8.28	0.92	0.39	1.79	0.76	0.47	0.71	1.01	1.80	0.21	1.13	1.15	NaN	0.41	2.21	0.57	1.20	0.46	NaN	NaN	0.48	1.28	NaN	NaN	0.42	0.64
MAP2K2	P36507	0.81	0.61	0.84	0.86	0.66	0.48	1.03	0.65	0.77	1.17	0.68	1.85	0.58	0.66	0.75	0.18	1.19	0.97	NaN	NaN	0.61	1.02	NaN	NaN	2.17	0.68
MAP4	E7EVA0	NaN	0.58	0.54	3.00	1.97	0.82	NaN	NaN	1.40	NaN	1.64	0.76	NaN	NaN	0.47	0.12	1.23	0.76	3.75	NaN	1.39	0.99	2.56	2.22	NaN	NaN
MAP4	P27816-5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MARK2	Q7KZ17-8	0.66	0.01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.72	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MAU2	Q9Y6X3	0.86	0.62	NaN	NaN	0.93	1.30	0.92	NaN	1.07	NaN	1.25	0.92	NaN	1.67	1.08	0.28	3.14	0.30	12.24	0.08	4.56	0.21	NaN	0.80	NaN	NaN
MAU2	Q9Y6X3-2	NaN	NaN	NaN	NaN	1.38	1.35	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.22	0.37	8.03	0.14	1.93	0.26	NaN	NaN	NaN	NaN	NaN
MAX	Q8TAX8	1.02	1.71	3.10	1.03	1.38	1.89	2.24	1.63	1.66	2.27	1.99	1.67	NaN	NaN	0.75	4.76	1.56	1.69	6.02	0.17	2.33	1.37	0.51	2.58	2.48	NaN
MBD2	Q9UBB5	0.38	0.79	1.48	0.71	0.97	1.04	1.00	1.02	1.11	0.89	1.06	0.87	0.33	2.55	0.68	0.32	0.65	1.46	15.14	0.01	0.86	1.17	6.56	0.08	2.02	0.54
MBTPS1	Q14703	NaN	NaN	NaN	NaN	0.35	0.49	NaN	NaN	NaN	NaN	0.32	NaN	NaN	NaN	NaN	NaN	0.61	NaN	NaN	NaN	0.27	0.62	NaN	NaN	0.69	0.69
MCRS1	Q96EZ8-3	0.58	0.33	2.31	0.79	1.08	1.26	1.35	1.15	0.89	1.62	1.46	1.22	0.26	5.42	0.37	0.53	0.80	1.57	7.29	0.18	2.37	0.62	1.71	1.12	NaN	0.49
MEA66	Q9HAF1-2	0.22	1.83	1.12	1.14	0.99	1.27	0.78	1.04	1.07	0.87	1.17	1.51	NaN	NaN	0.75	0.65	1.16	1.15	0.70	2.39	0.86	1.77	1.86	2.03	0.88	0.93
MECP2	P51608-2	1.04	1.09	0.50	0.93	0.55	0.91	0.58	0.70	0.88	0.65	0.75	0.71	0.62	0.78	0.60	5.77	0.62	0.98	5.15	1.0	0.77	0.74	3.74	0.32	0.58	0.58
MED10	Q9BTT4	0.67	NaN	NaN	NaN	0.89	0.54	0.60	NaN	0.60	1.10	0.83	0.76	NaN	NaN	0.22	NaN	0.52	0.90	0.22	NaN	1.25	0.53	NaN	NaN	NaN	NaN
MED16	Q9Y2X0-2	0.53	0.64	0.71	0.56	0.84	0.41	0.51	NaN	0.49	0.98	0.82	0.64	0.52	1.13	0.27	0.99	0.52	0.64	1.59	0.47	0.88	0.34	0.74	1.15	NaN	NaN
MIER1	Q8N108-19	NaN	NaN	0.62	0.66	0.72	1.49	0.80	NaN	0.87	1.05	NaN	NaN	NaN	NaN	NaN	0.84	2.15	0.31	2.07	0.09	2.39	0.32	NaN	NaN	NaN	NaN
MIER2	Q8N344	NaN	NaN	NaN	NaN	0.53	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.08	0.21	3.85	NaN	2.66	0.20	NaN	NaN	NaN	NaN
MIF	I4AY87	NaN	0.09	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.44	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MITF	A8K5K3	NaN	NaN	NaN	NaN	NaN	NaN	1.02	NaN	2.45	1.33	NaN	NaN	NaN	NaN	NaN	NaN	0.78	NaN	1.75	0.27	NaN	0.63	0.55	2.12	NaN	NaN
MLX	Q9UH92	0.44	0.72	1.46	0.37	0.99	1.11	1.27	0.86	0.72	1.21	0.98	0.86	0.38	2.25	0.24	0.31	0.77	0.93	10.58	0.21	1.34	0.62	0.25	2.64	2.25	0.34
MLXIPL	H7C1V3	NaN	0.58	1.29	0.42	0.75	0.85	NaN	NaN	NaN	NaN	0.72	0.53	NaN	NaN	NaN	NaN	0.68	0.75	NaN	0.12	0.90	0.80	NaN	NaN	NaN	NaN
MLXIPL	Q9NP71-3	0.31	0.55	1.38	0.28	0.94	1.04	0.94	0.65	0.60	1.17	0.87	0.80	0.46	2.00	0.22	0.21	0.71	0.83	10.76	0.17	1.11	0.74	0.27	1.85	NaN	0.44
MMS22L	E2QRD4	NaN	NaN	NaN	0.71	0.38	1.24	0.19	NaN	NaN	NaN	0.98	0.72	0.54	0.90	NaN	NaN	0.96	0.67	1.27	0.57	0.26	0.90	0.78	0.70	0.91	NaN
MNT	Q99583	0.43	1.20	1.60	0.82	1.14	1.74	1.11	1.11	1.18	1.49	1.19	1.27	0.74	2.24	0.37	6.72	1.26	1.41	3.72	0.23	1.23	1.13	0.33	2.83	2.65	0.51
MOB4	Q9Y3A3-3	0.48	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	32.32	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MORC2	Q9Y6X9-2	1.04	1.36	NaN	NaN	1.77	0.30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.08	0.67	NaN	0.35	1.07	1.16	NaN	0.63	1.27	1.22	NaN	NaN
MPHOSPH8	Q99549	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.07	NaN	NaN	12.07	NaN	3.00	0.28	NaN	0.08	2.63	0.22	NaN	NaN	NaN	NaN
MRPL13	Q9BYD1	0.49	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MTA1	Q13330	NaN	NaN	NaN	NaN	NaN	0.39	NaN	NaN	NaN	NaN	0.68	0.57	0.42	NaN	NaN	NaN	0.34	1.09	NaN	0.48	0.42	0.95	NaN	0.25	NaN	NaN
MTA1	Q13330-3	0.19	0.84	0.99	0.55	0.86	0.69	0.71	0.84	0.85	0.65	0.81	0.73	0.51	1.61	0.51	0.62	0.50	1.17	3.36	0.11	0.50	1.15	2.12	0.23	1.05	0.47
MTA2	Q94776	0.45	1.04	1.02	0.71	0.99	0.85	0.76	0.83	0.98	0.76	0.92	0.82	0.42	1.52	0.82	0.65	0.66	1.11	5.90	0.03	0.70	1.05	6.43	0.11	1.34	0.45
MTA3	E7EQY4	0.34	0.91	0.71	0.68	0.77	0.83	0.86	0.80	0.81	0.75	0.79	0.66	0.31	1.77	0.57	0.55	0.67	0.85	8.18	0.04	0.66	0.87	3.02	0.16	1.02	0.51
MTF2	Q9Y483	0.39	1.58	0.84	0.62	0.85	0.97	0.71	0.73	0.85	1.01	0.78	1.11	0.51	NaN	1.39	1.66	1.06	0.81	1.23	0.50	1.08	0.81	0.84	1.94	0.91	0.98
MYC	B4E1N7	NaN	NaN	3.87	0.90	1.95	2.08	2.31	NaN	1.61	1.76	2.18	2.17	NaN	NaN	NaN	NaN	1.68	1.93	4.29	0.23	2.20	1.41	0.79	1.61	NaN	NaN
MYL6B	P14649	NaN	5.52	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.64	1.31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NAA40	Q86UY6	0.80	0.62	0.80	0.84	0.91	0.95	0.91	0.72	0.80	0.98	0.87	0.84	0.57	1.21	0.89	1.07	0.90	0.77	0.97	0.90	0.89	0.77	0.89	0.97	0.35	2.29
NACA	E9PAV3	6.94	0.53	1.73	1.66	1.39	1.01	0.79	NaN	0.42	0.81	1.03	0.93	NaN	NaN	0.72	0.37	1.13	NaN	NaN	1.59	1.10	1.05	1.37	0.78	NaN	NaN
NAP1L1	F8VY35	1.04	2.68	2.06	0.90	2.30	0.84	1.09	1.74	1.85	1.18	0.76	1.98	0.80	1.90	1.95	1.35	2.12	0.71	1.19	1.22	1.85	0.81	2.19	0.45	1.20	1.86
NAP1L4	B7ZAK9	NaN	0.21	4.07	0.69	2.71	0.49	NaN	NaN	1.03	1.26	NaN	NaN	NaN	NaN	1.25	0.50	NaN	NaN	1.13	0.57	1.01	NaN	1.52	NaN	NaN	NaN
NBN	O60934	NaN	0.17	1.10	1.69	3.36	2.43	2.31	2.33	2.03	2.83	2.09	1.80	0.95	2.48	0.32	0.17	3.19	1.28	0.19	0.38	1.75	1.56	0.49</			

		H3Δ1-20_F	H3Δ1-20_R	H4K20me1_F	H4K20me1_R	H4K20me3_F	H4K20me3_R	H3Kc27me1_F	H3Kc27me1_R	H3Kc27me2_F	H3Kc27me2_R	H3Kc27me3_F	H3Kc27me3_R	H3K9me1_F	H3K9me1_R	H3K9me2_F	H3K9me2_R	H3K9me3_F	H3K9me3_R	H3K9me3 meCpG_F	H3K9me3 meCpG_R	H3K9me3 H4K20me3_F	H3K9me3 H4K20me3_R	meCpG_F	meCpG_R	H4R3me2_F	H4R3me2_R	
Gene names	Protein IDs																											
NFATC2IP	Q8NCF5	0.47	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.45	0.26	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NFIA	B1AKN8	0.60	0.38	0.78	0.49	0.41	1.62	NaN	NaN	0.21	0.53	0.84	0.68	0.24	2.93	0.09	2.99	0.18	2.64	69.81	0.15	2.04	0.25	1.04	0.87	NaN	0.26	
NFIA	Q12857-2	NaN	NaN	0.23	0.60	0.44	NaN	NaN	NaN	NaN	NaN	0.70	0.48	NaN	NaN	0.05	NaN	0.17	NaN	3.42	0.04	1.63	0.27	1.13	0.91	NaN	NaN	
NFIB	O00712-5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NFIB	Q5VW27	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.43	NaN	NaN	NaN	NaN	
NFIB	Q5VW28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NFIB	Q5VW30	NaN	0.37	0.98	0.59	0.43	1.74	NaN	0.64	0.43	NaN	0.83	0.73	0.28	6.29	0.10	7.91	0.19	2.70	17.59	0.08	1.90	0.31	1.07	0.91	NaN	NaN	
NFIC	P08651	0.73	0.85	0.97	0.56	0.40	1.54	0.26	0.27	0.19	0.42	0.84	0.75	0.11	7.60	0.11	11.19	0.19	2.33	16.73	0.01	2.02	0.27	1.41	0.99	NaN	0.20	
NFIC	P08651-5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.75	NaN	NaN	NaN	NaN	
NFIC	P08651-6	NaN	NaN	NaN	NaN	0.54	1.49	NaN	NaN	NaN	NaN	1.03	0.84	NaN	NaN	NaN	NaN	0.28	NaN	NaN	NaN	2.75	0.39	NaN	NaN	NaN	NaN	
NFIX	B4DM25	NaN	0.76	1.17	0.80	0.47	2.17	NaN	NaN	0.66	NaN	1.02	0.98	0.24	5.04	0.16	NaN	0.28	2.77	19.53	0.14	1.68	0.48	1.52	0.87	NaN	NaN	
NFIX	C9JWJ8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	6.98	0.63	1.73	0.44	NaN	NaN	NaN	NaN	
NFIX	D2DXM9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.24	NaN	NaN	NaN	NaN	NaN	NaN	
NFRKB	Q6P4R8-3	0.70	0.44	1.87	0.60	0.88	1.62	1.18	0.87	0.65	2.45	1.27	1.08	0.38	6.92	0.27	0.56	0.48	2.35	36.59	0.02	2.53	0.50	1.35	2.70	1.91	0.32	
NIPBL	Q6KC79	0.86	0.66	0.58	0.89	0.87	0.78	NaN	0.69	1.01	0.42	1.15	0.73	1.53	0.68	1.95	0.12	3.88	0.13	14.75	0.19	3.86	0.12	NaN	1.06	0.91	0.61	
NIPBL	Q6KC79-2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.16	NaN	NaN	NaN	NaN	2.31	0.07	NaN	NaN	NaN	NaN	
NOLC1	Q14978-2	0.60	0.61	0.68	1.21	0.91	0.76	0.98	1.23	1.02	0.81	0.65	1.34	0.89	0.73	0.83	0.66	1.21	0.56	1.12	2.64	0.37	1.85	2.18	1.32	0.44	2.62	
NOLC1	Q14978-3	NaN	NaN	0.90	1.22	1.21	1.11	1.17	1.66	1.38	1.17	0.81	1.48	1.15	1.11	NaN	NaN	1.38	0.69	1.21	3.02	0.46	2.39	2.42	1.51	0.38	3.33	
NPM1	E5R198	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NPM1	P06748	1.11	1.51	1.63	0.97	2.06	0.75	0.84	1.34	1.20	1.02	0.80	1.81	1.14	1.42	1.06	0.47	1.60	1.03	0.57	1.76	1.07	1.66	0.95	0.83	0.98	1.43	
NR2C1	H9NIM2	0.43	1.18	0.85	0.60	1.32	1.40	1.34	NaN	0.53	2.23	1.19	0.86	1.21	0.91	0.32	0.20	0.96	1.18	3.06	0.36	2.20	0.58	0.56	0.95	NaN	NaN	
NS5ATP4A	Q09GN0	0.27	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.49	0.18	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
ORC2	Q13416	0.66	0.73	1.03	1.03	7.63	0.04	1.18	0.88	0.84	0.94	2.93	0.20	1.07	0.89	0.76	0.19	2.86	0.19	3.24	0.78	7.26	0.08	2.47	1.05	0.82	0.95	
ORC3	Q9UBD5-2	0.82	0.83	0.89	0.96	6.95	0.04	1.00	0.87	0.83	0.81	2.39	0.19	1.06	0.79	0.88	0.44	2.41	0.18	2.98	0.83	5.33	0.06	2.79	1.10	0.75	0.97	
ORC3L	Q9UBD5	NaN	NaN	1.85	NaN	6.65	0.16	NaN	0.68	1.47	1.45	2.26	0.24	NaN	NaN	0.68	0.10	NaN	0.24	NaN	1.29	6.24	0.25	NaN	NaN	NaN	0.75	
ORC5	A4D0P7	0.59	0.74	1.25	1.14	7.95	0.08	1.48	1.05	0.99	1.07	2.88	0.21	1.00	0.96	0.64	0.41	2.66	0.24	6.95	0.40	6.65	0.08	2.84	1.07	1.34	1.27	
ORC6	B3KMP9	0.67	0.60	NaN	NaN	4.67	0.16	1.20	NaN	NaN	0.82	0.97	0.58	NaN	0.59	0.50	0.07	1.70	0.34	1.09	1.06	NaN	0.49	NaN	NaN	NaN	NaN	
OXSRI	Q95747	1.88	1.50	NaN	NaN	0.66	0.50	0.82	0.86	0.82	1.30	0.69	1.45	0.66	1.07	1.51	0.25	1.21	1.16	NaN	NaN	0.64	0.85	NaN	NaN	1.51	0.70	
PARD3	Q8TEW0-11	1.90	NaN	NaN	NaN	NaN	NaN	0.41	0.49	0.33	NaN	NaN	NaN	NaN	NaN	0.10	0.16	NaN	NaN	1.24	0.25	1.08	NaN	0.48	1.02	NaN	0.79	
PARK7	Q99497	NaN	0.12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.16	NaN	NaN	NaN	0.21	NaN	NaN	NaN	0.84	0.39	NaN	NaN	
PATZ1	Q59H11	1.01	0.50	0.92	1.34	0.54	1.78	NaN	NaN	1.55	1.18	0.73	0.71	NaN	NaN	0.57	3.04	0.43	2.06	2.87	0.62	1.10	0.73	1.03	0.90	3.95	NaN	
PAXIP1	B4DEQ6	0.63	1.15	2.01	0.54	1.06	1.16	NaN	NaN	0.50	1.89	1.20	1.37	0.90	1.72	0.43	0.07	0.51	2.13	6.18	0.25	1.71	0.88	2.02	1.34	0.67	NaN	
PBX2	P40425	NaN	NaN	NaN	NaN	2.33	0.51	NaN	NaN	0.64	NaN	1.25	0.38	NaN	NaN	0.60	3.30	0.93	0.55	12.87	0.10	1.67	0.45	1.23	0.50	NaN	NaN	
PC4	Q6IBA2	1.04	1.15	7.12	0.83	0.84	0.95	0.98	0.96	0.99	0.92	0.88	0.84	0.87	0.88	0.96	1.43	0.77	0.94	0.98	1.27	0.72	0.95	1.07	1.00	0.87	1.08	
PCMT1	H7BY58	1.18	1.40	0.39	0.50	0.38	0.44	0.33	0.23	0.38	0.35	0.54	1.25	NaN	NaN	1.22	4.14	0.21	0.67	0.38	0.36	0.26	0.46	0.49	0.37	0.76	0.32	
PCNA	P12004	0.38	0.61	1.09	0.63	0.68	1.36	0.80	1.07	0.50	1.68	0.85	0.96	0.34	1.78	0.23	1.30	0.45	1.70	3.63	0.09	2.09	0.35	0.93	1.06	1.28	0.41	
PCSK9	Q8NBP7	NaN	NaN	NaN	0.45	0.46	0.51	NaN	NaN	0.38	0.44	0.55	NaN	0.62	0.33	NaN	0.77	0.40	0.48	NaN	NaN	0.32	0.40	NaN	NaN	NaN	NaN	
PDIA5	Q14554	NaN	NaN	NaN	NaN	3.57	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
PES1	B5MCF9	NaN	1.93	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.25	0.71	NaN	NaN	1.92	NaN	NaN	NaN	1.16	1.42	NaN	NaN	1.21	NaN	NaN	NaN	
PFN1	P07737	0.20	0.24	NaN	NaN	NaN	NaN	0.35	1.42	NaN	0.34	NaN	NaN	NaN	NaN	NaN	NaN	0.85	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.15	NaN	
PGAM1	Q6P6D7	0.30	0.16	NaN	NaN	NaN	0.88	0.37	2.87	0.31	0.51	NaN	NaN	NaN	NaN	0.24	0.25	NaN	2.91	0.58	1.33	NaN	NaN	0.26	NaN	NaN	NaN	
PGBD3	A8K4Q3	NaN	NaN	0.99	0.52	0.79	0.63	0.68	0.59	0.78	0.55	0.73	0.79	0.31	1.99	NaN	NaN	0.49	0.86	9.50	0.09	0.84	0.65	4.32	0.19	0.85	0.24	
PGK1	P00558	0.43	0.12	2.36	NaN	NaN	0.72	NaN	4.63	2.16	0.30	NaN	NaN	NaN	NaN	0.53	0.19	1.02	0.61	0.32	2.11	0.39	NaN	1.10	0.22	NaN	NaN	
PHF1	E9PQT8	NaN	NaN	NaN	NaN	7.63	0.23	NaN	NaN	5.32	0.34	9.40	0.21	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.81	NaN	NaN	NaN	NaN	
PHF14	Q94880	0.07	2.12	1.45	1.13	1.46	1.30	1.00	1.26	1.03	1.51	1.08	1.29	0.67	3.93	0.50	0.95	1.25	1.24	1.23	2.00	1.11	1.43	1.27				

Gene names	Protein IDs	H3Δ1-20_F	H3Δ1-20_R	H4K20me1_F	H4K20me1_R	H4K20me3_F	H4K20me3_R	H3K9c27me1_F	H3K9c27me1_R	H3K9c27me2_F	H3K9c27me2_R	H3K9c27me3_F	H3K9c27me3_R	H3K9me1_F	H3K9me1_R	H3K9me2_F	H3K9me2_R	H3K9me3_F	H3K9me3_R	H3K9me3 meCpG_F	H3K9me3 meCpG_R	H3K9me3 H4K20me3_F	H3K9me3 H4K20me3_R	meCpG_F	meCpG_R	H4R3me2_F	H4R3me2_R
PNP	Q8N7G1	NaN	0.18	NaN	NaN	NaN	NaN	0.39	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.47	0.59	NaN	NaN	0.27	NaN	NaN	NaN	1.01	NaN	NaN	NaN
POF1B	Q8WVV4	NaN	NaN	NaN	NaN	NaN	NaN	0.12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.15	NaN	NaN	NaN	0.21	NaN	NaN	NaN	NaN	NaN	NaN	NaN
POGZ	Q7Z3K3-5	NaN	0.69	1.51	0.93	1.07	1.35	0.86	NaN	0.59	1.84	1.13	1.16	2.13	1.51	0.77	NaN	5.91	0.11	47.32	0.14	7.48	0.08	5.82	0.79	NaN	NaN
POGZ	Q7Z3K3-7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.77	NaN	15.19	NaN	NaN	NaN	6.14	NaN	NaN	NaN
POLD3	B7ZAQ5	0.51	NaN	NaN	0.59	0.42	0.48	NaN	NaN	NaN	0.84	0.77	0.66	NaN	0.40	0.70	0.45	NaN	NaN	NaN	1.00	0.58	0.53	1.72	NaN	NaN	0.54
POLG	E5KNX5	NaN	NaN	NaN	0.41	0.38	0.38	0.29	NaN	0.73	1.00	0.37	0.51	NaN	NaN	0.45	NaN	0.19	0.52	NaN	NaN	0.49	0.47	NaN	NaN	NaN	0.42
POLR1A	B7ZKR9	NaN	NaN	NaN	NaN	NaN	NaN	1.13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.76	0.34	NaN	NaN	0.37	NaN	NaN	0.51	NaN	NaN	NaN	NaN
POLR1A	O95602	1.85	2.21	1.02	1.08	0.47	1.98	1.03	0.76	0.39	3.52	1.80	0.75	0.76	1.40	1.15	2.91	0.21	2.19	27.77	0.24	1.94	0.41	0.73	2.40	0.75	0.95
POLR1A	B7ZKR9	NaN	NaN	NaN	NaN	NaN	NaN	1.13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.76	0.34	NaN	NaN	0.37	NaN	0.51	NaN	NaN	NaN	NaN
POLR1A	O95602	1.85	2.21	1.02	1.08	0.47	1.98	1.03	0.76	0.39	3.52	1.80	0.75	0.76	1.40	1.15	2.91	0.21	2.19	27.77	0.24	1.94	0.41	0.73	2.40	0.75	0.95
POLR1B	Q9H9Y6	1.94	2.83	1.08	1.06	0.42	2.20	1.21	0.82	0.38	3.98	2.30	1.06	0.40	3.60	1.22	3.63	0.25	2.48	16.97	0.15	2.54	0.54	0.75	2.26	NaN	NaN
POLR1C	Q96HT3	1.60	1.26	0.70	0.69	0.34	0.94	0.74	0.81	0.47	0.89	0.86	0.68	0.16	2.47	0.70	3.79	0.15	1.70	21.83	0.04	1.54	0.28	0.79	3.44	5.07	0.17
POLR1D	Q9Y2S0	NaN	0.82	NaN	NaN	0.56	1.12	0.87	0.89	0.56	0.87	0.92	0.76	NaN	NaN	0.48	2.13	0.24	1.80	0.05	NaN	1.97	0.36	NaN	NaN	NaN	NaN
POLR1E	Q9GZS1-2	NaN	1.67	1.01	1.06	0.51	0.97	1.28	0.89	0.50	3.10	1.75	0.69	0.26	NaN	1.14	NaN	0.31	1.10	5.58	0.12	1.78	0.36	0.74	NaN	NaN	NaN
POLR2E	E5KT65	1.06	0.96	0.74	0.69	0.38	1.00	0.74	0.85	0.47	0.94	0.93	0.70	0.21	2.04	0.39	3.56	0.17	1.63	15.20	0.06	1.67	0.30	0.73	1.89	3.37	0.16
POLR2H	C9JLU1	1.41	1.40	0.67	1.01	0.46	0.98	0.83	0.96	0.52	0.93	1.00	0.80	0.20	2.22	0.36	1.39	0.17	1.61	15.42	NaN	1.70	0.30	NaN	NaN	NaN	NaN
POLR2K	P53803	NaN	NaN	NaN	NaN	0.45	1.17	0.78	0.95	0.57	1.01	1.03	NaN	NaN	NaN	NaN	0.20	1.99	NaN	NaN	1.73	0.39	NaN	NaN	NaN	NaN	NaN
POLR2L	P62875	1.01	0.92	NaN	NaN	0.37	1.01	0.77	0.87	0.55	1.11	0.82	NaN	NaN	NaN	0.45	1.29	0.15	1.88	NaN	NaN	1.85	0.30	0.77	NaN	NaN	NaN
POLR3A	O14802	1.03	0.93	0.36	0.44	0.37	0.54	0.26	0.66	0.45	0.67	0.52	0.52	0.26	1.80	0.44	2.46	0.13	1.25	24.78	0.15	1.15	0.20	1.05	3.00	2.76	0.18
POLR3B	Q7Z3R8	1.16	1.02	0.42	0.54	0.36	0.56	0.27	0.67	0.44	0.71	0.51	0.67	0.28	NaN	0.41	1.78	0.14	1.37	4.80	0.21	1.29	0.26	1.19	2.45	NaN	0.25
POLR3C	Q9BU14	0.98	0.91	0.32	0.52	0.36	0.56	0.31	0.71	0.47	0.60	0.50	0.51	0.35	0.78	0.53	1.53	0.14	1.37	9.50	0.34	1.17	0.22	1.04	2.03	2.96	0.21
POLR3D	P05423	0.34	0.50	0.45	0.46	0.34	0.58	0.28	0.71	0.46	0.67	0.56	0.54	0.22	1.67	0.34	2.68	0.13	1.21	14.87	NaN	1.16	0.23	NaN	2.19	NaN	0.11
POLR3E	Q9NVU0	0.94	0.83	0.30	0.37	0.38	0.54	0.26	0.66	0.45	0.45	0.51	0.53	0.35	1.59	0.41	3.68	0.15	1.21	1.97	0.20	1.06	0.18	1.24	NaN	NaN	0.20
POLR3F	Q53F18	1.43	0.82	0.36	0.47	0.31	0.58	0.30	0.79	0.45	0.73	0.45	0.53	NaN	NaN	0.37	5.03	0.13	1.41	NaN	0.18	1.25	0.21	0.79	3.01	2.57	0.21
POLR3G	D6R9U7	1.56	0.86	0.35	0.40	0.34	0.47	0.34	0.77	0.46	0.73	0.56	0.53	NaN	NaN	0.39	1.90	0.18	1.52	11.83	0.32	1.39	0.21	NaN	2.53	NaN	0.17
POLR3GL	A6NGX6	NaN	NaN	0.29	NaN	0.29	0.52	0.28	0.62	0.44	0.66	0.54	0.59	NaN	NaN	0.51	1.17	0.19	1.32	NaN	NaN	1.12	0.22	NaN	NaN	NaN	NaN
POLR3H	E7ER22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.85	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
POLR3H	Q9Y535	1.52	0.67	0.33	0.45	0.35	0.64	0.33	0.76	0.50	0.75	0.60	0.56	0.15	1.68	0.36	1.62	0.14	1.52	26.33	0.26	1.39	0.20	NaN	3.06	NaN	0.10
POLR3H	E7ER22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.85	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
POLR3H	Q9Y535	1.52	0.67	0.33	0.45	0.35	0.64	0.33	0.76	0.50	0.75	0.60	0.56	0.15	1.68	0.36	1.62	0.14	1.52	26.33	0.26	1.39	0.20	NaN	3.06	NaN	0.10
POLR3K	Q9Y2Y1	NaN	NaN	NaN	NaN	0.44	0.89	0.37	1.14	0.65	0.90	0.64	0.84	NaN	NaN	NaN	1.62	0.24	1.66	NaN	NaN	1.61	0.27	NaN	NaN	NaN	NaN
POM121C	A8CG34	0.51	NaN	NaN	NaN	NaN	NaN	0.04	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
POP4	O95707	0.47	NaN	NaN	NaN	NaN	1.06	NaN	NaN	NaN	NaN	0.92	0.93	0.79	0.80	0.52	0.24	0.93	NaN	1.53	1.37	1.28	1.03	0.78	NaN	NaN	NaN
POU2F1	H0YLB5	NaN	0.16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PPIA	A8K486	0.16	0.13	1.79	1.62	1.53	1.02	0.31	NaN	1.40	0.38	1.05	0.94	0.60	0.79	0.21	0.34	0.90	0.64	0.43	NaN	0.73	0.57	0.79	0.47	NaN	NaN
PRC1	H9KV59	0.70	NaN	NaN	NaN	0.54	0.19	NaN	NaN	NaN	NaN	0.33	NaN	NaN	NaN	0.43	0.52	0.22	NaN	0.58	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PRC1	O43663	0.57	0.79	0.47	0.36	0.72	0.29	0.46	1.30	0.49	0.45	0.63	0.30	0.35	1.35	0.60	0.48	0.80	0.25	1.90	0.32	0.75	0.26	0.42	NaN	0.46	0.26
PRDM10	B7ZL72	NaN	NaN	NaN	0.49	0.65	0.95	NaN	NaN	0.21	1.87	0.67	0.73	0.57	NaN	0.23	NaN	0.41	0.80	98.83	0.13	1.91	0.23	2.67	1.16	NaN	NaN
PRDX1	Q06830	0.65	0.32	1.55	1.73	1.64	0.83	0.33	1.87	0.69	0.49	0.79	0.85	0.82	1.13	0.61	1.07	1.31	1.04	0.43	1.01	0.93	0.82	1.00	0.52	0.87	0.57
PRDX2	P32119	0.41	0.16	1.50	NaN	NaN	0.88	0.07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.39	0.70	NaN	NaN	0.21	0.95	0.81	NaN	0.88	0.41	NaN	NaN
PRDX6	P30041	0.47	0.22	2.87	2.76	1.07	0.91	0.42	2.67	1.61	0.50	0.74	NaN	0.99	0.87	0.40	0.54	1.16	1.16	0.28	1.77	0.65	0.59	1.13	0.30	1.65	0.64
PRIC295	E1NZA1	1.50	1.05	0.51	0.84	0.70	0.29	1.24	0.36	0.47	0.65	0.48	0.54	0.99	0.64	1.11	0.45	1.02	0.64	0.46	0.39	0.41	0.77	NaN	NaN	1.05	0.44
PRKD2	Q8NCK8	0.19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PRMT5	O14744	1.00	0.82	0.84	0.60	0.72	0.85	NaN	NaN	0.60	0.85	0.90	0.71	2.91	0.28	1.18	0.49	0.33	1.52	1.05	0.81	0.95	0.51	1.15	0.89	NaN	0.28
PRPF40A	O75400-2	2.16	1.94	0.38	0.83	0.95	0.47	1.09	0.96	NaN	0.35	1.53	0.66	NaN	0.50	1.07	1.25	1.02	0.39	1.67	NaN	0.46	0.96	1.15	NaN	0.30	0.75
PRPF40A	O75400-3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.92	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PSAT1	B4DHQ3	NaN	0.11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.28	1.81	NaN	NaN	0.94	NaN	NaN	NaN
PSMA6	G3V5Z7	0.21	0.18	NaN	NaN	NaN	NaN	0.28	2.39	NaN	0.63	0.50	NaN	NaN	NaN	0.20	0.04	1.13	0.82	0.49	1.28	1.19	1.07	1.12	0.74	NaN	NaN
PSMA7	O14818	0.61	0.30	1.21	NaN	1.02	1.15	0.60	1.40	0.40	0.61	0.99	0.70	1.10	0.69	0.50	0.17	0.71	NaN	0.57	1.07	0.70	0.88	1.06	0.87	NaN	NaN
PSMD3	Q6IBN0	0.85	0.64	2.35	0.37	1.07	1.08	NaN	NaN	0.71	0.79	NaN	0.84	NaN	2.68	0.76	0.37	1.23	1.12	0.87	0.06	NaN	1.02	NaN	1.02	NaN	NaN
PUM1	Q14671	0.76	4.67	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.90	0.29	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PUM1	Q5T1Z4	0.76	3.89	1.60	NaN	1.30	1.26	1.40	NaN	0.72	0.95	NaN	NaN	1.65	0.67	0.92	0.09	1.28	1.34	NaN	NaN	0.49	1.35	NaN	NaN	1.95	0.97
PURA	Q2NLD4	1.73	0.83	0.47	3.16	0.60	2.00	NaN	NaN	NaN	NaN	NaN	1.29	NaN	NaN	0.90	2.93	0.66	1.64	3.50	0.51	1.65					

Gene names	Protein IDs	H3Δ1-20_F	H3Δ1-20_R	H4K20me1_F	H4K20me1_R	H4K20me3_F	H4K20me3_R	H3Kc27me1_F	H3Kc27me1_R	H3Kc27me2_F	H3Kc27me2_R	H3Kc27me3_F	H3Kc27me3_R	H3K9me1_F	H3K9me1_R	H3K9me2_F	H3K9me2_R	H3K9me3_F	H3K9me3_R	H3K9me3 meCpG_F	H3K9me3 meCpG_R	H3K9me20me3_F	H3K9me20me3_R	meCpG_F	meCpG_R	H4R3me2_F	H4R3me2_R
RAI14	Q9P0K7-2	NaN	3.57	1.03	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.19	NaN	0.49	NaN	NaN	NaN	0.19	NaN	NaN	NaN	NaN
RALGAPA2	Q2PPJ7-3	0.67	NaN	NaN	1.31	NaN	NaN	NaN	NaN	NaN	NaN	0.88	NaN	NaN	NaN	1.34	0.14	NaN	NaN	NaN	NaN	NaN	35.51	0.19	NaN	NaN	NaN
RAVER1	E9PAU2	0.88	1.48	0.53	0.73	1.03	0.48	1.11	0.42	0.60	0.73	0.49	0.57	1.19	0.56	0.82	0.43	1.09	0.75	0.75	NaN	0.67	0.96	NaN	NaN	1.29	0.50
RBBP4	Q09028-3	0.38	0.67	1.15	0.76	0.99	1.10	0.89	0.95	0.88	1.02	0.99	0.87	0.42	2.03	0.71	2.17	1.11	0.82	4.30	0.08	1.48	0.60	3.37	0.27	1.64	0.55
RBBP7	Q16576-2	0.34	0.69	1.16	0.78	0.92	1.01	0.95	0.95	0.98	0.89	1.03	0.87	0.40	2.26	0.64	1.61	0.77	1.00	12.27	0.03	0.87	0.95	3.85	0.17	1.66	0.56
RBBP7	Q6FHQ0	0.33	0.65	1.28	0.76	0.93	1.01	1.00	1.01	1.10	0.89	1.03	0.85	0.37	2.62	0.66	3.20	0.73	1.14	20.37	0.02	0.80	1.16	5.14	0.16	1.87	0.55
RBM4	A8K9U0	1.64	0.98	0.68	1.27	1.02	1.15	0.87	0.92	NaN	0.89	0.95	0.89	NaN	0.81	1.17	0.55	NaN	0.73	5.52	0.58	1.10	0.96	7.67	0.72	0.84	0.91
RBM7	J3KPD3	0.83	0.63	1.19	1.61	0.95	0.91	1.11	1.08	1.03	1.06	NaN	1.37	0.72	1.22	0.79	0.74	1.08	NaN	NaN	NaN	1.16	1.11	1.51	NaN	NaN	NaN
RBM7	Q6IRX3	0.57	0.55	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
RECQL	P46063	0.48	0.81	0.33	0.54	0.79	0.68	0.60	NaN	0.51	0.57	0.56	0.43	0.25	1.24	0.62	2.59	0.38	0.81	1.45	0.59	1.11	0.26	1.50	1.15	0.55	0.34
RFX1	P22670	0.71	0.87	0.80	0.89	0.43	1.18	NaN	NaN	0.62	NaN	0.90	0.84	0.84	0.59	0.70	0.08	0.45	1.89	11.12	0.14	1.54	0.50	2.16	0.55	NaN	0.53
RFXANK	O14593	0.54	1.26	1.03	0.30	0.83	0.58	NaN	NaN	0.34	1.10	0.58	0.55	0.25	1.55	0.27	0.43	0.57	0.56	20.68	0.05	1.34	0.26	1.58	0.59	NaN	NaN
RIOK3	B4E1Q4	0.50	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.43	0.15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
RMND1	Q5SZ82	1.11	NaN	4.02	NaN	NaN	NaN	0.90	NaN	NaN	0.80	0.79	0.92	NaN	NaN	NaN	NaN	0.70	0.90	0.92	NaN	0.71	NaN	NaN	NaN	NaN	NaN
RNF213	Q63HN8	0.78	0.42	0.72	0.98	NaN	0.17	NaN	NaN	NaN	0.45	0.74	0.57	1.13	0.42	0.54	0.23	1.28	0.27	NaN	0.32	NaN	0.79	NaN	NaN	1.54	0.41
RNPS1	H3BV80	5.57	1.30	0.48	1.49	0.79	0.55	0.91	1.12	2.23	0.39	1.34	1.17	NaN	0.59	1.79	2.26	1.04	0.57	1.08	2.62	0.37	1.71	NaN	NaN	0.48	0.54
RPL7L1	R4GMU7	NaN	NaN	0.37	1.78	0.95	0.70	NaN	NaN	1.72	NaN	0.98	0.62	NaN	0.88	0.86	0.35	1.00	0.50	0.85	0.72	0.60	1.07	1.09	0.66	0.67	1.03
RPLP2	P05387	1.30	1.51	NaN	NaN	1.30	0.93	0.93	1.10	1.44	1.16	0.71	NaN	NaN	1.01	1.12	0.39	1.02	1.08	1.06	37.48	1.08	1.50	0.66	0.43	NaN	0.72
RPRC1	D3DPS3	1.19	1.71	1.37	0.27	NaN	0.60	NaN	NaN	NaN	0.58	0.43	1.08	NaN	NaN	1.43	1.00	NaN	NaN	0.86	0.67	1.47	0.36	0.85	0.78	NaN	NaN
RPRD1B	Q9NQG5	0.69	0.54	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.80	0.81	0.66	NaN	1.03	NaN	1.58	1.04	1.12	NaN	1.90	1.04	NaN
RSF1	Q96T23-2	NaN	0.60	0.16	0.62	0.41	0.45	NaN	0.48	0.54	0.39	0.35	0.36	0.55	NaN	0.56	NaN	0.42	0.53	0.43	1.43	0.38	0.71	0.07	0.02	0.03	0.36
RSL1D1	Q76021	0.64	2.86	0.17	2.10	0.58	0.75	0.57	NaN	1.12	0.40	0.75	0.45	0.70	0.62	2.07	0.60	0.57	0.53	1.15	NaN	0.53	0.60	1.68	0.66	0.26	0.88
RSRC2	Q7L412-2	0.60	0.80	2.06	0.22	0.66	0.56	1.47	0.58	0.50	0.35	NaN	NaN	NaN	NaN	0.56	0.93	0.55	NaN	0.68	1.02	NaN	0.70	1.05	0.78	0.37	NaN
S100A11	P31949	NaN	NaN	NaN	NaN	NaN	NaN	0.17	NaN	NaN	NaN	NaN	0.70	NaN	NaN	NaN	0.39	NaN	NaN	NaN	NaN	NaN	NaN	0.34	NaN	NaN	NaN
S100A14	Q9HCY8	NaN	NaN	NaN	NaN	NaN	NaN	0.01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SAP130	H7BXF5	0.55	1.15	1.26	1.26	0.77	2.28	1.47	1.31	1.72	1.35	1.24	1.10	0.60	2.95	0.40	NaN	1.01	1.19	4.46	0.29	0.88	1.47	0.76	4.12	1.49	0.63
SAP18	Q00422	11.09	1.94	0.33	1.48	0.76	0.40	0.74	0.86	2.32	0.32	NaN	NaN	NaN	NaN	3.14	2.03	1.17	0.50	NaN	NaN	0.54	1.25	1.70	1.73	0.42	NaN
SAP30	Q75446	0.71	0.79	1.47	1.38	0.83	3.28	1.73	1.41	1.63	1.92	1.62	1.22	0.64	2.90	0.54	0.82	1.29	1.60	10.80	0.19	1.23	1.77	0.81	4.83	3.45	0.87
SARS	Q53HA4	NaN	0.24	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.50	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SCML2	B4DZR9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SCML2	Q9UQR0	0.63	1.53	8.39	0.10	1.02	0.82	NaN	NaN	0.96	0.81	0.73	1.13	0.65	0.72	4.14	0.42	2.39	0.26	7.89	0.16	1.43	0.45	16.71	0.03	0.22	2.16
SCYL2	Q6P3W7	1.60	1.60	0.63	0.71	1.04	0.60	0.89	0.26	0.88	1.32	0.62	NaN	1.10	0.65	1.01	0.14	1.16	0.88	NaN	NaN	0.58	1.14	NaN	NaN	1.91	0.63
SDHB	Q0QEY7	NaN	NaN	NaN	0.50	0.43	0.43	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.25	NaN	NaN	NaN	NaN	NaN
SEMG1	P04279-2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.09	NaN	NaN
SEMG2	A8K6Z6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.31	NaN	NaN
SEPT10	E7EW69	NaN	NaN	NaN	NaN	1.52	0.34	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.96	NaN	NaN	1.62	1.89	NaN	NaN	0.23	1.77	NaN	NaN	NaN	NaN
SEPT2	Q15019	1.17	1.78	0.69	0.81	1.69	0.31	0.44	0.67	0.88	NaN	0.58	0.62	0.73	0.51	0.76	0.23	2.03	0.13	0.58	0.54	0.11	2.30	0.81	0.89	0.49	0.44
SEPT6	Q8NFH9	NaN	NaN	0.65	1.18	1.89	0.35	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.33	0.23	NaN	NaN	0.40	2.72	NaN	NaN	NaN	NaN	NaN
SEPT7	Q16181	1.54	1.59	NaN	NaN	1.67	0.30	NaN	NaN	NaN	NaN	0.55	0.40	NaN	0.62	NaN	NaN	2.08	0.11	0.61	NaN	0.14	2.00	NaN	NaN	NaN	NaN
SEPT7	A8K3D0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SEPT8	B7ZVZ1	NaN	NaN	NaN	NaN	1.72	0.34	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.32	NaN	NaN	NaN	0.36	NaN	NaN	NaN	NaN	NaN
SEPT9	Q9UHD8	1.29	2.84	0.51	0.63	1.59	0.24	NaN	NaN	NaN	NaN	0.59	0.63	0.97	0.26	0.91	0.26	1.56	0.12	NaN	NaN	0.14	1.56	NaN	NaN	0.57	1.00
SERPINB3	P29508	NaN	NaN	NaN	NaN	NaN	NaN	0.20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.30	NaN	NaN	NaN	NaN	0.54	NaN	NaN
SETX	Q7Z333	0.72	0.83	1.13	0.99	1.27	0.47	0.14	1.02	1.01	1.13	1.19	1.14	1.00	0.74	0.71	0.06	0.89	0.94	0.04	NaN	0.58	0.63	NaN	NaN	1.21	0.31
SF3A1	Q15459	2.02	1.00	0.32	2.80	1.58	0.46	0.68	1.07	1.71	0.58	0.98	1.20	1.96	0.61	1.20	1.05	1.82	0.49	0.68	1.49	0.35	2.11	NaN	0.88	0.57	1.22
SF3A3	B3KY12	1.43	1.01	0.33	2.09	1.32	0.43	0.55	1.05	1.63	0.60	0.65	0.85	1.01	0.80	1.18	1.90	1.58	0.41	0.77	0.93	0.31	2.03	0.97	0.87	0.30	1.13
SF3B1	Q75533	1.53	1.09	0.26	2.11	1.54	0.37	0.51	1.00	1.53	0.43	0.86	1.05	2.12	0.66	1.25	1.22	1.50	0.38	1.29	0.49	0.32	1.81	1.06	1.12	0.27	1.08
SF3B2	E9PJ04	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.83	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SF3B4	Q53FG6	1.34	0.70	0.47	NaN	1.17	0.47	0.77	NaN	1.81	0.55	1.00	1.31	1.95	0.82	1.24	2.04	1.54	0.49	1.13	1.05	0.30	1.92	1.08	0.87	0.47	1.12
SFN	P31947	0.31	0.14	NaN	NaN	NaN	0.40	0.16	NaN	NaN	0.27	NaN	NaN	NaN	NaN	0.25	NaN	0.65	NaN	NaN	NaN	NaN	0.40	0.13	NaN	NaN	NaN
SHCBP1	B2RDX0	NaN	1.95	NaN	NaN	1.20	0.42	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.15	NaN	NaN	0.37	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SHKBP1	B2R6W9	NaN	2.05	2.61	247.82	NaN	NaN	NaN	NaN	18.49	NaN	8.02	NaN	7.50	NaN	NaN	NaN	14.85	10.88	NaN	NaN	NaN	7.93	NaN	12.58	9.65	2.47
SHPRH	Q149N8	0.54	0.39	0.55	1.15	1.13	0.80	NaN	NaN	2.46	0.28	NaN	NaN	0.99	0.71	0.04	2.02	1.25	0.51	0.78	1.63	0.31	1.80	NaN	NaN	1.07	0.59
SIN3A	Q96ST3	0.60	1.10	1.14	1.43	0.78	2.21	1.22	1.28	1.33	1.39	1															

Gene names	Protein IDs	H3Δ1-20_F	H3Δ1-20_R	H4K20me1_F	H4K20me1_R	H4K20me3_F	H4K20me3_R	H3Kc27me1_F	H3Kc27me1_R	H3Kc27me2_F	H3Kc27me2_R	H3Kc27me3_F	H3Kc27me3_R	H3K9me1_F	H3K9me1_R	H3K9me2_F	H3K9me2_R	H3K9me3_F	H3K9me3_R	H3K9me3 meCpG_F	H3K9me3 meCpG_R	H3K9me3 H4K20me3_F	H3K9me3 H4K20me3_R	meCpG_F	meCpG_R	H4R3me2_F	H4R3me2_R	
SMCHD1	A6NHR9	0.54	0.91	0.61	0.97	0.99	0.73	0.58	NaN	0.79	0.65	1.02	0.73	3.26	0.67	1.44	0.17	2.21	0.24	47.25	0.19	3.01	0.19	NaN	1.06	NaN	NaN	
SMCHD1	L0R6P7	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.78	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
SNC73	Q9UP60	NaN	NaN	NaN	NaN	47.70	NaN	0.06	NaN	NaN	NaN	0.05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.12	
SNRPA1	P09661	1.80	0.90	0.39	2.21	1.43	0.47	0.65	1.06	1.57	0.58	0.90	0.99	1.68	0.56	1.10	0.99	1.57	0.43	0.94	1.41	0.38	1.79	1.17	1.01	0.44	1.17	
SNRPB2	P08579	1.62	0.96	0.37	1.84	1.40	0.47	0.69	1.03	1.53	0.48	0.77	0.89	1.68	0.57	1.06	1.03	1.53	0.43	0.98	1.26	0.41	1.65	1.08	1.07	0.49	1.14	
SNRPD3	B4DJP7	1.72	0.87	0.55	2.08	1.24	0.62	0.80	1.11	1.71	0.71	1.10	1.14	NaN	0.67	1.13	1.02	1.43	0.64	0.80	NaN	0.72	1.38	0.73	0.79	0.59	1.22	
SP1	G5E9M8	0.84	NaN	NaN	0.54	NaN	1.23	NaN	NaN	0.30	1.72	0.83	0.61	NaN	NaN	0.47	0.49	NaN	1.74	1.34	NaN	2.52	0.24	NaN	NaN	NaN	NaN	
SP3	Q8WUU3	0.70	NaN	0.79	0.55	0.79	1.58	0.96	1.01	0.45	1.88	0.94	0.81	NaN	NaN	0.39	3.36	0.43	2.14	1.40	0.82	3.09	0.28	1.10	0.70	2.61	0.28	
SPATA5L1	Q9BVQ7	1.54	1.33	1.50	NaN	0.89	0.60	NaN	NaN	0.97	NaN	0.71	1.46	1.15	0.86	1.11	0.45	0.88	1.20	0.91	NaN	0.82	0.78	NaN	NaN	NaN	NaN	
SPIN1	Q9Y657	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.88	0.22	3.36	0.38	7.35	0.13	NaN	NaN	NaN	NaN	
SPIN2B	Q5JZB8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	7.57	NaN	12.49	0.13	9.55	0.12	NaN	NaN	NaN	NaN	
SPRR1A	B7ZLF8	NaN	NaN	0.04	NaN	NaN	NaN	0.06	NaN	NaN	0.03	0.14	0.18	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
SPRR1B	P22528	NaN	NaN	NaN	NaN	NaN	NaN	0.05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
SPTY2D1	Q68D10-2	1.07	1.61	0.58	0.62	0.53	1.01	0.49	1.21	1.18	0.77	0.71	0.91	2.42	NaN	1.21	0.62	0.55	0.89	0.65	0.69	0.76	0.52	0.86	0.47	0.67	1.30	
SREK1	B3KRJ9	1.25	0.69	2.82	0.21	1.15	0.53	NaN	NaN	NaN	NaN	NaN	NaN	2.27	0.79	0.96	0.83	1.22	0.54	NaN	NaN	0.57	1.04	NaN	NaN	NaN	0.45	0.70
SRRM1	A9Z1X7	7.44	3.28	0.27	0.72	0.35	0.21	0.34	0.48	0.72	0.15	0.45	0.39	NaN	0.27	1.64	1.39	0.37	0.22	0.78	1.27	0.34	0.42	NaN	NaN	0.20	0.27	
SRSF6	A8K588	7.33	1.79	0.52	2.00	0.92	0.75	0.97	1.25	2.36	0.31	1.51	1.23	2.06	1.16	1.53	1.63	1.32	0.59	1.63	1.50	0.77	1.38	NaN	0.96	0.33	0.63	
SUDS3	Q9H7L9	0.60	0.77	1.45	1.34	0.72	3.05	1.47	1.43	1.51	1.69	1.52	1.10	0.58	3.30	0.41	3.57	1.19	1.47	9.63	0.24	1.11	1.76	0.71	3.53	2.46	1.07	
SUPT3H	Q75486	0.64	0.52	1.02	0.55	0.45	0.97	0.84	NaN	0.58	0.99	0.82	0.77	1.55	0.41	0.44	0.45	0.47	1.02	NaN	0.09	1.53	0.38	0.58	2.32	1.71	0.40	
SUZ12	Q15022	0.35	2.04	0.76	0.65	0.89	0.93	0.76	0.98	0.89	0.89	0.92	0.84	0.61	1.37	1.44	1.45	1.33	0.50	1.80	0.89	1.16	0.44	1.07	1.72	1.05	0.67	
SV2A	A8K6Q3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.03	NaN	NaN	NaN	
SV2A	L8E840	NaN	NaN	NaN	0.96	NaN	NaN	NaN	0.83	NaN	NaN	NaN	1.01	NaN	NaN	NaN	NaN	1.06	1.22	NaN	NaN	0.85	NaN	NaN	NaN	0.83	0.84	
SVL	Q95425-2	NaN	4.06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.08	0.62	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
SYAP1	B2RBI2	NaN	0.12	0.88	1.23	NaN	0.17	0.82	NaN	NaN	0.78	NaN	NaN	NaN	NaN	NaN	NaN	1.36	0.90	NaN	NaN	NaN	1.09	NaN	NaN	1.30	0.55	
SYDE1	B2RD93	1.00	NaN	2.41	NaN	3.42	0.36	NaN	NaN	NaN	NaN	0.56	1.46	NaN	NaN	0.68	0.13	1.17	0.62	1.37	0.75	0.80	1.05	NaN	NaN	NaN	NaN	
SYN1	P17600	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.02	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.18	NaN	NaN	NaN	
SYT1	J3KQA0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.14	NaN	NaN	NaN	
TADA1	Q96BN2	0.57	0.57	0.75	0.54	0.49	1.11	0.75	NaN	0.54	0.95	0.81	0.72	NaN	NaN	0.55	0.30	0.51	0.97	5.32	0.17	1.31	0.40	0.77	1.95	0.33	NaN	
TADA2B	Q86TJ2	NaN	NaN	0.96	0.53	0.49	1.02	0.12	NaN	0.59	0.86	0.94	0.72	0.29	2.07	NaN	NaN	0.44	0.96	0.12	0.12	1.39	0.38	NaN	1.07	0.16	NaN	
TADA3	Q75528	NaN	0.30	1.76	0.83	1.15	1.15	1.20	0.99	0.89	1.45	1.10	0.94	NaN	NaN	0.29	1.42	0.94	1.13	1.85	0.76	1.40	0.79	1.22	1.29	1.57	0.92	
TAF1	P21675-6	NaN	NaN	0.53	0.62	0.51	NaN	0.71	NaN	0.48	0.83	0.67	0.79	NaN	NaN	0.18	NaN	0.29	1.34	4.80	0.28	1.47	0.32	NaN	NaN	NaN	NaN	
TAF11	Q15544	NaN	NaN	NaN	NaN	0.71	NaN	0.41	1.26	0.61	0.94	0.74	1.04	NaN	NaN	NaN	NaN	0.36	0.52	NaN	NaN	1.76	0.42	NaN	0.80	NaN	NaN	
TAF1A	Q15573	NaN	NaN	0.46	NaN	0.27	0.96	0.60	0.70	0.42	0.94	0.78	0.67	NaN	NaN	NaN	NaN	0.22	1.33	7.79	0.15	1.02	0.48	0.41	1.36	NaN	NaN	
TAF1B	Q53T94	NaN	NaN	NaN	1.00	0.84	NaN	0.53	0.78	NaN	NaN	0.99	0.75	NaN	NaN	NaN	NaN	0.29	NaN	NaN	0.22	NaN	0.40	0.48	NaN	NaN	0.98	
TAF1D	Q9H5J8	NaN	NaN	1.91	0.73	NaN	NaN	0.72	0.67	0.35	1.03	0.83	0.75	NaN	NaN	0.52	NaN	3.21	1.75	9.84	0.22	0.95	0.51	0.51	2.33	NaN	NaN	
TAF5	Q15542	0.24	1.04	0.44	0.75	0.63	1.10	0.36	0.91	0.49	0.98	0.77	0.87	NaN	2.06	0.15	1.30	0.31	1.40	2.51	0.15	1.50	0.37	1.34	2.18	NaN	NaN	
TAF6	J3KR72	0.33	0.90	0.41	0.67	0.63	0.83	0.30	0.88	0.45	0.79	0.58	0.88	0.35	1.13	0.14	3.94	0.32	1.43	2.56	0.45	1.33	0.31	NaN	2.36	NaN	0.50	
TALDO1	P37837	NaN	0.08	NaN	NaN	1.18	NaN	NaN	NaN	NaN	0.37	NaN	NaN	NaN	NaN	NaN	NaN	1.82	0.67	0.55	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
TARS	Q53GX7	0.46	2.76	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.54	0.42	1.07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
TBRG1	Q3YBR2	NaN	NaN	2.66	0.70	0.86	2.03	3.27	NaN	1.06	2.91	1.70	1.24	0.34	4.99	NaN	1.32	0.49	2.50	6.11	0.10	2.92	0.64	NaN	1.46	NaN	NaN	
TCHH	Q07283	NaN	NaN	NaN	NaN	2.83	0.54	NaN	NaN	NaN	NaN	NaN	0.78	NaN	NaN	NaN	0.86	3.06	0.20	NaN	NaN	1.80	0.34	NaN	NaN	0.15	NaN	
TCOF1	E7ETY2	NaN	NaN	NaN	NaN	0.74	1.46	2.08	1.40	0.67	1.54	2.41	0.62	NaN	NaN	NaN	NaN	0.43	1.06	NaN	NaN	1.73	0.52	NaN	NaN	NaN	NaN	
TCOF1	E9PHK9	0.94	1.11	0.68	0.93	0.67	0.91	1.34	1.04	0.62	1.36	1.73	0.66	0.32	0.94	0.51	3.86	0.40	0.85	2.11	0.63	1.41	0.47	0.84	1.00	0.80	0.42	
TCOF1	Q13428-3	NaN	NaN	0.62	0.79	0.70	0.98	1.18	0.90	0.58	1.07	1.51	0.62	NaN	NaN	NaN	NaN	5.82	0.33	0.79	NaN	NaN	2.02	0.47	NaN	NaN	1.08	0.25
TF	Q53H26	NaN	NaN	NaN	NaN	NaN	NaN	0.11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
TFAP4	Q6FHM5	0.99	0.42	2.57	1.24	1.09	3.69	1.73	1.59	1.26	2.30	1.93	1.79	NaN	NaN	0.08	3.25	0.84	3.84	26.96	0.06	3.54	1.12	2.42	3.00	1.64	NaN	
TFPT	G5E9B5	0.65	0.36	2.24	0.60	0.64	1.64	1.41	0.91	0.61	2.70	1.34	1.23	0.19	5.94	0.29	1.41	0.50	2.37	NaN	0.06	3.41	0.52	0.99	2.26	4.32	0.42	
TGM1	B0AZN7	NaN	NaN	NaN	NaN	NaN	NaN	0.02	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
TGM3	Q08188	NaN	NaN	0.23	NaN	NaN	NaN	0.18	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.08	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
TIAL1	Q01085-2	1.75	1.20	1.07	1.49	1.11	0.98	NaN	NaN	NaN	NaN	1.41	1.34	1.37	1.06	1.37	0.49	NaN	2.35	NaN	1.72	1.15	1.11	NaN	NaN	1.03	1.29	
TIAL1	Q2TSD2	NaN	1.01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
TIMP2	P16035	NaN	2.25	1.29	1.13	NaN	1.65	1.49	1.35	NaN	1.83	1.32	1.39	0.96	NaN	NaN	NaN	NaN	1.19	1.67	NaN	NaN	NaN	1.58	NaN	NaN	NaN	
TINF2	B4DFJ1	NaN	NaN	1.16	1.30	1.16	2.40	NaN	NaN	NaN	2.61	NaN	1.55	NaN	NaN	NaN	NaN	0.62	2.35	2.44	1.02	1.66	1.50	1.60	2.38	NaN	NaN	
TKT	P29401	0.16	0.10	2.42	NaN	NaN	0.44	0.22	1.70	1.37	0.30	NaN	NaN	NaN	NaN	0.23	0.54	0.89	0.63	0.19	2.15	0.62	NaN	0.85	NaN	NaN	NaN	
TM																												

Gene names	Protein IDs	H3Δ1- 20_F	H3Δ1- 20_R	H4K20 me1_F	H4K20 me1_R	H4K20 me3_F	H4K20 me3_R	H3Kc27 me1_F	H3Kc27 me1_R	H3Kc27 me2_F	H3Kc27 me2_R	H3Kc27 me3_F	H3Kc27 me3_R	H3K9m e1_F	H3K9m e1_R	H3K9m e2_F	H3K9m e2_R	H3K9m e3_F	H3K9m e3_R	H3K9m e3 meC pG_F	H3K9m e3 meC pG_R	H3K9m e3 H4K 20me3_F	H3K9m e3 H4K 20me3_R	meCpG _F	meCpG _R	H4R3m e2_F	H4R3m e2_R
ZHX1	Q9UKY1	0.35	0.71	0.74	0.48	2.25	0.23	NaN	NaN	0.46	0.62	0.49	0.52	NaN	NaN	0.29	NaN	1.85	0.18	41.17	0.11	0.47	0.64	24.81	0.12	NaN	NaN
ZHX2	Q9Y6X8	0.42	0.57	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.48	NaN	NaN	NaN	NaN	0.25	0.56	0.42	0.71	33.67	0.12	0.96	0.34	26.09	0.09	NaN	NaN
ZHX3	A8K8Q0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	19.41	0.14	NaN	0.39	17.43	0.15	NaN	NaN
ZKSCAN3	Q9BRR0	0.52	0.46	0.53	0.92	0.62	0.85	0.67	NaN	0.60	0.69	0.68	0.53	0.53	0.88	0.53	NaN	0.55	0.73	4.22	0.50	0.59	0.59	0.84	0.83	NaN	0.50
ZMYM1	Q5SVZ6	0.88	1.22	1.07	NaN	0.62	2.22	1.05	1.55	0.57	2.61	1.30	1.96	NaN	1.79	0.50	0.31	0.30	3.05	6.90	3.06	2.12	0.88	1.18	8.86	NaN	NaN
ZMYM2	Q9UBW7	0.76	0.63	NaN	NaN	0.99	NaN	NaN	NaN	0.63	0.69	0.70	0.57	NaN	NaN	0.53	0.19	1.64	0.30	26.45	0.16	2.82	0.24	0.90	NaN	NaN	NaN
ZMYND8	Q9ULU4-11	NaN	0.88	0.33	0.51	0.74	0.58	NaN	NaN	0.61	0.43	0.72	0.58	0.47	2.29	0.46	NaN	0.49	0.90	15.95	0.26	1.15	0.38	10.81	0.90	NaN	NaN
ZMYND8	Q9ULU4-9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	7.83	NaN	NaN	NaN
ZNF131	P52739	0.44	1.21	1.17	0.70	1.12	1.07	0.96	0.92	0.84	1.02	0.97	1.02	0.58	1.36	0.48	3.26	1.03	0.81	1.63	0.85	0.92	0.85	0.59	2.17	0.91	0.74
ZNF148	Q9UQR1	1.40	0.65	1.36	0.81	0.57	2.07	0.98	1.11	1.04	1.41	1.17	1.17	0.32	4.18	0.25	3.30	0.42	2.43	20.20	0.18	1.77	0.73	1.68	0.81	2.40	0.35
ZNF280D	A8K124	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.72	2.82	0.45	8.04	0.19	2.19	0.15	NaN	NaN	NaN	NaN
ZNF282	Q86YG2	NaN	NaN	1.77	0.77	NaN	NaN	NaN	NaN	0.69	2.37	1.12	1.14	NaN	NaN	NaN	NaN	0.44	3.37	NaN	NaN	2.22	0.66	NaN	NaN	NaN	NaN
ZNF354A	Q86Y64	NaN	NaN	0.43	NaN	0.54	0.92	NaN	NaN	NaN	NaN	0.71	0.62	0.41	0.88	NaN	NaN	0.48	0.64	NaN	NaN	0.74	0.56	NaN	NaN	NaN	NaN
ZNF444	Q53F81	1.06	0.29	0.81	1.16	0.60	2.13	1.09	0.95	0.82	0.92	0.98	0.83	0.59	1.41	0.41	0.69	0.38	1.79	2.50	0.39	1.10	0.67	1.24	0.93	NaN	NaN
ZNF451	Q4KMR5	1.51	0.13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.45	NaN	NaN	NaN	0.29	0.41	NaN	NaN	NaN	NaN
ZNF580	Q9UK33	NaN	NaN	2.87	0.63	1.17	1.34	1.37	NaN	0.96	1.29	1.18	1.13	NaN	NaN	NaN	0.18	2.17	0.68	5.07	0.15	3.22	0.55	1.86	2.04	NaN	0.52
ZNF598	Q86UK7-2	1.10	1.44	NaN	NaN	3.62	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.24	0.55	0.12	NaN	1.19	1.03	0.65	NaN	NaN	1.00	NaN	NaN	NaN
ZNF629	Q9UEG4	0.50	0.46	NaN	NaN	NaN	NaN	0.37	NaN	0.44	NaN	0.88	0.41	NaN	0.99	0.31	NaN	1.16	0.33	2.98	0.17	4.22	0.24	NaN	NaN	NaN	NaN
ZNF644	Q9H582	0.41	0.47	1.92	1.27	1.12	2.01	1.28	1.38	1.35	1.86	1.68	1.43	2.46	1.46	0.55	0.20	0.65	2.98	13.72	0.16	1.65	1.29	2.56	1.72	1.87	0.58
ZNF687	Q8N1G0	NaN	NaN	0.51	0.48	0.59	0.68	NaN	NaN	0.43	0.64	0.66	0.49	0.60	NaN	0.55	0.72	0.39	0.75	46.11	0.23	0.86	0.36	11.67	0.86	NaN	NaN
ZNF770	A8K5X3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.42	NaN	0.40	0.43	NaN	NaN	NaN	NaN	NaN	NaN	0.51	0.57	NaN	NaN	NaN	NaN	NaN	NaN
ZRANB2	Q95218-2	1.16	1.14	4.46	0.13	NaN	NaN	0.71	0.69	0.64	0.51	NaN	NaN	NaN	NaN	1.12	2.62	1.18	0.64	0.69	1.20	0.50	0.71	NaN	0.88	0.47	0.65
ZSCAN20	P17040-3	NaN	1.13	1.05	0.71	0.80	1.08	NaN	NaN	NaN	1.26	1.06	1.06	NaN	NaN	0.29	NaN	1.00	1.08	6.98	0.17	1.26	0.84	0.44	2.04	15.72	NaN

8 Curriculum Vitae

Personal data

Name: Zimmermann, Nadin
Date of birth: 20th January 1984
Place of birth: Berlin, Germany
Nationality: German

Education

04.2012 – 06.2016	GGNB Biomolecules: Structure – Function - Dynamics, Göttingen PhD thesis in Chromatin Biochemistry Group, Max Planck Institute for Biophysical Chemistry, Göttingen “Systematic analysis of heterochromatin modification readout”
10.2011	Diploma in Biology
10.2005 – 10.2011	Martin-Luther-University, Halle-Wittenberg Diploma thesis in Department of Developmental Genetics, Martin-Luther-University, Halle-Wittenberg “Reinigung differentieller dLSD1 Proteinkomplexe und Charakterisierung ihrer Funktion beim Silencing transponibler Elemente in der Keimbahn und im Soma von <i>Drosophila melanogaster</i> ”
06.2003	Abitur
08.1996 – 06.2003	High School “Jules-Verne-Oberschule”, Berlin, Germany

Stipends

02.2012 – 06.2016	Max Planck fellowship
-------------------	-----------------------